Do Algorithms Dream Up Electric Terrorists?

The Israeli Security Agency's Use of Predictive AI in Counter-Terrorism (2015-2017)

By

Adv. Nery Ramati LLM

A Thesis submitted for the degree of PhD
School of Law and Government
Dublin City University

Supervisor: Professor Maura Conway

December 2025

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original and have conformed to the regulations on the use and declaration of Generative AI, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.
I hereby certify that no Generative Artificial Intelligence (Gen AI) tools have been used in the creation of the thesis.

Signed: Nery Ramati

ID No. 20214474

Date: 15.12.2025

Acknowledgments

Making a list of acknowledgements for a PhD thesis is a strange thing, especially when you write on a subject that was already such a part of your life for many years before starting to write the thesis. Where do you start with the thanks? Is starting with the primary school teacher who unjustly punished you and thereby developed your sense of justice going too far? In any case, I would like to apologise to the many people I cannot mention here due to the rules. Thank you!

I would like to start, of course, with huge thanks to my supervisor, Prof. Maura Conway, for her guidance, insights and inspiration. This thesis was written during turbulent times, and her calm approach and her understanding on what kind of support I needed during each period of the research was remarkable. I would like to thank and acknowledge the financial support of the Irish Research Council through the Andrew Grene Postgraduate Scholarship in Conflict Resolution Award. I would also to thank the whole academic and administrative team of Dublin City University for their support, with special thanks to Prof. Paula Rivetti, who was of enormous help with her clear, wise and straightforward remarks throughout the research. Special thanks to my friend, Adv. Karin Torn Hibler, for her friendship and tenacity in leading the legal case. I would also like to thank my interviewees for having the courage to speak to me although, many times, we were on opposite sides. I would like to thank all of the brave Palestinian and Israeli human rights lawyers and academics who helped me with good advice and especially Gaby Lasky, Hasan Jabarin, Sharon Weill, Neta Ziv, Orna Ben Naftali, Fadi Qawasmi, Michael Sfard, Smadar ben Natan and Hedi Viterbo. And, of course, last but not least, to my amazing wife who was my unofficial second supervisor and to my amazing daughters, who had to put on a brave face when I was rambling on about my research.

Table of Contents

List of Figures

List of Tables

Do Algorithms Dream Up Electric Terrorists?: The Israeli Security Agency's Use of Predictive AI in Counter-Terrorism (2015-2017). By Nery Ramati

Abstract

Counter-terrorism agencies increasingly leverage artificial intelligence in their operations, with a strong focus on predictive tools that can identify potential individual attackers or attacks before they occur. To be effective, these tools need to analyse open-source and private surveillance data continuously. Despite advancements, there are significant concerns about their effectiveness and the potential legal and societal impacts. While the broader implications of AI technology have been widely discussed in academia, the specific use of preventive person-based AI tools in counter-terrorism remains underexplored.

This research examines the AI tool employed by the Israeli Security Agency (ISA) in the Occupied Palestinian Territories from 2015 to 2017 to identify individual attackers before they strike. Through interviews, open-source data, freedom of information requests, and extensive legal proceedings, the study investigates the tool's creation, design, capabilities, effectiveness, and impact. It proposes methods to enhance the tool's efficiency and reduce potential harm. Utilising critical terrorism studies, the research questions the necessity of such tools even under optimal conditions and suggests alternative applications of AI to mitigate political violence. The research also suggests that at least part of the lure of those AI tools is found in the lessening of moral responsibility they offer counter-terrorism agencies and agents, as difficult moral decisions are allocated to algorithms. This study not only fills a gap in academic literature but is also relevant for AI developers, law enforcement and policymakers.

Chapter 1 – Introduction

This research had its genesis in a very specific date and place. It was on 25 February 2016 in a small prefab building named 'Courtroom Number 7' of the Israeli military courts[1] near Ramallah in the Occupied Palestinian Territories (OPT). The room was used for remand requests by the Israeli military prosecution and was, as usual, packed with Palestinian detainees in brown prison uniforms (separated from their anxious families, who were trying to speak to them, by Israeli prison guards), prosecutors, (seated and surrounded by piles of files and trying to understand which case was ready to be presented), and defence lawyers, including me, who had to navigate between the detainees and the families whilst trying to get answers from the prosecutors regarding the charges. There was also Rafiq, a 21-year-old Druze soldier whose official role was as a translator but, in his unofficial capacity, was managing the whole court room, and a reserve judge, Major Shlomo Katz, a real estate lawyer from Tel Aviv who was called to reserve duty as a military judge and was trying his best, not always successfully, not to fall asleep.

It was Thursday afternoon, the last working day of the court that week, and everyone was in a rush to finish the case files. The Israeli military courts, with their 99.7% conviction rate, had a repeating ritual when it came to remand requests. The prosecution would declare that it had filed an indictment against the defendant and give the file with the indictment and the evidence to the defence attorney who would see them for the first time. The defence attorney would browse the file for a few seconds, nod to the prosecution, and

[1] A detailed description of how the Israeli military legal system works can be found in Chapter Six.

declare that the sides had agreed to postpone the hearing for a month in order to allow the defence to study the materials. Rafiq, the translator, would cough lightly near the judge's ear and he would wake up immediately and dictate a decision to postpone the hearing and keep the defendant in custody until that hearing. And so, one by one, they passed in front of the judge, young people suspected of stone throwing, family men who had been caught working in Israel without a valid permit, students who had joined a student body that had been declared illegal, a butcher who had been carrying his butcher's knife with him when stopped at a random check point, and a driver that had been caught driving a stolen car. Each had their own story, a worried family and a prosecution file with witnesses, reports, testimonies and exhibits.

On that day, I was representing eight young people from the village of Nebi Salah. They had all been charged with being involved in stone throwing against the Israeli Army during the weekly demonstrations in the village. I managed to finish the hearings for four of them but was stuck in court because the prison guards had not yet brought the other four from the nearby prison. Sitting on the bench waiting for my turn, I was approached by a family I did not know. The father explained that his 19-year-old son, Sohaib, had been arrested at his home in the middle of the night five days previously and they had been told that he had a court appearance on that day and asked me if I could represent him as he was just entering the court room. I went to talk with the son.

Sohaib was very shy, almost not communicative at all. The only thing I managed to understand from him was that, after his arrest, he had been questioned only once by the police. He was asked if he had been planning an attack against Israelis and he denied this.

He was shown his Facebook page and was asked if it was his and he confirmed it. I asked him if the police had confronted him with a testimony of someone else saying he was planning an attack and he said no. This scenario seemed a bit strange to me. I had had my share of cases where people had been charged with planning an attack based on the testimony of another person, but then they had usually been confronted with this testimony, or any other evidence that existed. I had also had my share of cases where the army charged prominent Palestinian social media influencers with 'incitement' based on their Facebook activity. However, those cases had always been high profile and had involved a lot of media attention and social media buzz. I hadn't heard a thing about this timid young man who had already been under arrest for five days. I looked at Sohaib's Facebook profile and it confirmed my suspicion, he had only forty Facebook friends, most of which, according to their names, were his family. His posts were a mixture of pictures of luxury sports cars and pictures of recent attackers who had been part of the wave of attacks that started in 2015, with text praising their heroism, standard for almost any young Palestinian social media feed at that time. My conclusion, after seeing his Facebook page, was that the case was based on someone else's testimony, where it was claimed that he had planned an attack, and that the Facebook page was just a non-relevant extra.

As the hearing started, I was surprised to see that I was wrong. The indictment charged Sohaib with an incitement felony and there was no mention of any planning or attempt to attack. Only a list of five Facebook posts praising the dead attackers, and the likes each got, which were between five and twenty likes for each post. The indictment also had only one witness on the witness list, the policeman who had investigated Sohaib. When the

case started and I received Sohaib's file, I became even more confused. Sohaib had indeed been questioned by the police regarding planning an attack, although there was no evidence in the file for such a plan. The only evidence against him was his unimpressive Facebook page. Rafiq, the translator, looked at me, waiting for me to declare that I was asking for a postponement, but I decided to challenge the case on the spot. There was no real evidence to go over or to study, and the file looked like just an error. 'Your honour', I said, 'I think this case is one big mistake, and I would like to argue for my client's release'. Judge Katz woke up again and stared at me with astonishment. 'You are going to argue here, on a Thursday afternoon, when I have thirty more cases to go over?' He picked up the file and looked at it. 'What seems to be the problem?' 'My client is charged with incitement although almost no one saw his posts, and he was questioned about planning an attack despite there being no evidence for it.' Judge Katz picked up the file again and browsed the indictment. 'Ahh yes, it's an artificial intelligence case. I already had a few like this, this week.' He handed me back the file and was ready to rule.

'A what case?' I asked, astonished. 'An artificial intelligence case. The ISA now uses artificial intelligence to find who could be the next terrorist.' A piercing stare from the prosecutor, who stood up quickly, made the judge realise he had said much more than he should have. 'Anyway, it's not relevant now,' said the judge quickly. 'The decision on whether the defendant's posts are inciting or not is a question for the trial judge. My role is to check if he is dangerous and I rule that, under the current circumstances, he cannot be released on bail. Next case please!' It was then that I first learned the astonishing and almost unbelievable fact which would become publicly acknowledged a couple of years

later, the ISA has been using an AI preventive tool to try and predict who the next attacker will be, and they have been using this output to send people to jail based on its recommendations. This is where this research begins.

The above pages were written by me in 2020, at the beginning of my research. I wanted to document those events while they were still fresh in my head and I thought they could be nice opening paragraphs for the thesis. For me, in 2020, the idea of using artificial intelligence, as it was then, for such complex work was borderline science fiction, and the idea that Israel would use this technology to imprison people dystopian. It is widely known that, because the world is advancing at a very fast pace, research PhDs on current issues are aging quickly, no matter what the topic is. However, I think it is very hard to find two issues that have changed more dramatically over the last few years than the capabilities of artificial intelligence technology, and the Israeli/Palestinian conflict, making what in 2020 looked like a dystopian science fiction story, look like just a regular Monday now.

The 'leap forward' in the ability of AI technologies since 2020 has not so much been a leap but part of a continuous technological advancement in both hardware and software abilities. However, many attribute the appearance of the 'transformer model' in 2017 (Vaswani 2017) to the rapid development of the Large Language Model (LLM) which has made AI abilities much more accessible to the public in the shape of products such as chat bots (Luitse and Denkena 2021). In 2024, Chat GPT alone had 200 million weekly active users (Backlinko 2024), very different from what we knew about AI in 2016 when I first heard about the ISA using it, or even in 2020 when this research began.

The Israeli/Palestinian conflict in 2020 looked like a repetitive cycle of escalations that the world was losing interest in, with the Israeli government advancing a policy of trying to maintain the conflict at a low level, with short rounds of violence in Gaza, whilst continuing and advancing its settlement policies (Kingsley and Bergman 2021). The surprise attack by Hamas on 7 October 2023 has changed the course of this conflict forever. On the Israeli side, it was the deadliest day in the history of Israel, with more than a thousand civilians and soldiers dead, and tens of thousands displaced from their ruined houses. On the Palestinian side, the Israeli response led to the deadliest period in their history since the 1948 war, with more than fifty thousand people dead, mostly civilians, and hundreds of thousands of displaced people living in tents with an ever-growing humanitarian crisis. It has brought the conflict to the verge of an all-out regional war, with Lebanon, Yemen and Iran taking an active part in it (Abrams et al. 2024). Stating today that Israel was using AI technology to predict who the next attacker would be and that it was imprisoning Palestinians based on that, is therefore much less shocking than it was when this research began, especially since we now know, for example, that, during the Israel/Hamas war, Israel used AI technology to recommend targets to bomb in Gaza while having it calculate how many civilians might get hurt (Avraham 2023).

Having said all that, the fact that AI technology has advanced and continues to advance at a frantic pace since the case study on AI was launched, and the fact that the Israeli/Palestinian conflict has reached its bloodiest stage since the establishment of Israel, does not mean in any way that the research and its outputs have become irrelevant. The only thing this research might have lost, due to those changes, is the sensational nature of the case study. In many ways, the

tectonic changes, both in AI technology and in the region, have made the research much more relevant and needed.

The journey of completing this research was a rocky one, as it started hand in hand with the Covid-19 pandemic outbreak, which was handled very differently in Ireland, where I was based, and in Israel, my case study country. During my journeys back and forth, I managed to very narrowly escape prison time in Israel for attempting to leave the country before the end of my entering curfew, and being locked in an expensive hotel for two weeks, at my own expense, upon returning to Ireland. Obtaining the data was also not easy, as the Israeli army refused to provide all of the data I asked for in my freedom of information request, which led to a rare evidence-hearing case in the Israeli District Court (Case no. 55892-10-21 Ramati Vs Israel Defence Forces – Jerusalem District Court), which I won, and an even rarer appeal by the state to the Supreme Court (Case no. 637/23Israel Defence Forces Vs Ramati, Israeli Supreme Court), which I won again. To add to all of those, the horrible war between Israel and Hamas that broke out in October 2023, has not only shaken me personally but has also closed any window to obtain any further data from the Israeli authorities, as their sole focus is on the continuing fighting. Under those circumstances, between Covid, court hearings and war, I feel very lucky in having managed to complete this work.

Relevance and Potential Significance of the Research

This research attempts to answer a very clear research question: Could we and should we use predictive artificial intelligence (AI) in counter-terrorism to identify potential individual attackers before they attack? I sought to answer this question via a case study of an AI tool used by the

Israeli Security Agency (ISA) against Palestinians in the Occupied Palestinian Territories (OPT) between the years 2015 and 2017.

The growing availability of artificial intelligence technology has led to counter-terrorism agencies, and security forces in general, being interested in using its powers for their purposes, especially using its predictive powers to pre-identify attacks and attackers (Wall 2024). That need has grown exponentially, as it seems that more and more individuals have engaged in violence against crowds without any clear organisational support, a phenomenon that is known by many names: lone wolf attacks, school shootings, mass shootings, leaderless attacks, and more (Smith et al. 2015). The belief that predictive AI could be the solution in identifying these attackers might have originally come about as a result of what looked like a synchronised campaign by several Israeli security personnel who gave interviews and published articles (in a non-characteristic way, given the usual secrecy that surrounded them) about how an Israeli predictive AI tool had managed to identify potential individual Palestinian attackers prior to an attack and had successfully suppressed a wave of individual attacks, beginning in October 2015 (Y. 2018; Berbing 2019; Shragai 2019; Shahaf 2020; Y. S. 2021). This rush by security and law enforcement agencies towards technologies that predicted behaviours did not go unnoticed in academic circles and it was criticised from several different angles: this technology is not efficient and cannot predict such unique human behaviour (T. Munk 2019); this technology is not ready yet and is missing a lot of the protective elements that would make it truly effective (Ganor 2019); this technology is biased by nature as it is based on biased data and therefore should not be used (Babuta 2019; 2020); and this technology requires constant surveillance that does not meet legal requirements (Wall 2024). As security agencies are, by their nature, not very generous with their data, most of

these comments lacked an in-depth analysis of the tools used by those agencies, the data fed into them, how it was collected, what the exact output was, and what was done with it. But more than that, those articles were very much focused on the tool's current lack of effectiveness, with almost no theoretical discussion on the meaning of having and using such an effective tool in the sphere of counter-terrorism. This research fills those gaps, both practically and theoretically, while being aware of the obvious gap between them. On the practical side, this research presents the story of the ISA's predictive AI tool from beginning to end, following the logic of a counter-terrorism agency. It looks at the reasons why the ISA had a sudden need for such a tool, the creation of the tool, the activation of it, and its direct impact. The research assesses the tool effectiveness and looks into the theoretical aspect of using such a tool, whether the tool could be effective, the meaning behind creating such an effective tool, the reasons such tools are so sought after, the societal cost of operating it, and whether we can suggest an alternative, more ethical way to use predictive AI in order to reduce individual violent attacks. On the theoretical side, this research brings together concepts from critical terrorism studies and especially critical counter-terrorism studies, questioning some of the basic definitions used by counter-terrorism agencies when it comes to individual attacks, terrorism, violence, counter-terrorism and effectiveness, and joining them with the ever-growing body of critical thinking about AI and the concepts of inherent bias and the delegation of moral duties to algorithms. By discussing both the practical sides of activating such a tool and the theoretical questions that arise from it, this research tries to come up with a theoretical framework that also has a clear practical view regarding the use of predictive AI in attempting to predict individual violence.

Individual Attacks or Attackers in this Research

This research examines the ISA response to a phenomenon that was unknown to the ISA prior to it starting in October 2015: violent attacks by young Palestinians with no organisational affiliation, using knives, rifles, and cars in order to hurt and kill Israeli soldiers and civilians. The new aspect of those attacks was the lack of any organisational affiliation among the attackers. They were not recruited by anyone nor had they conspired with anyone prior to their attack. This phenomenon is, as mentioned above, known by many names and has been frustrating law enforcement and counter-terrorism agencies all over the world. For purposes of this research, I have chosen to refer to these as 'individual attacks' and 'individual attackers' as I have found that the more common definition of 'lone actor terrorism' is a wrong fit for the case study. The use of the word 'lone' derives from the original use of 'lone wolf terrorists'. The 'lone wolf' is a term that has been highly criticised in literature: 'it is misleading, it valorizes the perpetrator and it has been misused' (Berntzen and Bjørgo 2021, 133). The 'lone' part of it was described as applying a normative label to the attackers on a range from 'losers' to 'deranged' (Spaaij and Hamm 2015). As for the terms 'terrorists' or 'terrorism', I chose not to use them when describing that kind of specific attacks or attackers for a combination of reasons. First, the term terrorism, as will be discussed later at length, is an unstable social construct that is used to describe a wide range of actions and sometimes not used to describe very similar actions. In the US, the phenomenon of school shooters and school shootings is rarely named 'terrorism'. However, the characteristics of those attacks are in many ways similar to 'lone actor terrorism' as we know it (Gill et al. 2021). In regard to the specific case study, the question of the term 'terrorism' is becoming even more complex. First, many of the attacks that were part of the case study were carried out by

Palestinians in the occupied territory against Israeli military personnel and can be considered by some as part of the ongoing military conflict. Second, the words 'terrorist' and 'terrorism' are almost absent from the legal vocabulary that is relevant to the case study as, in the Occupied Palestinian Territories, the Israeli military law that applies refers to those as 'security offences' or 'security offenders'. To add to that, the Israeli military courts that were dealing with those cases chose to use the term individual attacks in their jurisprudence when describing such attackers. To conclude this point, the terms 'individual attacks' or 'individual attackers' describe both the lack of affiliation of the perpetrators and the action they perform, which is a violent attack carried out by a lone individual. In so doing, I avoid the difficulties that arise from using the term 'lone actor terrorism' and it fits the vocabulary that is used in the case study.

Scope of the Research and Limitations

The research will focus on the particular use of AI by security agencies to predict future individual violent human behaviour. This could be, as in the ISA case study, an attempt to discover who the next individual violent attacker will be. It does not dwell on the more general questions that arise from the use of algorithms to predict individual human behaviour, such as algorithms that predict what you might want to buy based on your behavioural patterns or algorithms that try to predict what you are searching for online, based on your interactions with them, or, more specifically, the role of algorithms in online (violent) radicalisation processes, on which there is a growing literature (Gill et al. 2017; Scrivens and Davies 2018; Whittaker and Charmin 2021) . There are two main differences that make the security agencies' specific use of predictive algorithms stand apart. The first difference is that the individual behaviour those agencies are looking for is usually very unique and, the more unique the behaviour, the less powerful the tools become (Fuchs et

al. 2023). While algorithms are very good at predicting that you really feel like ordering Chinese food now, based on your previous behaviour, you and a few hundred thousand people on the planet at the same time, it will have a harder time guessing that you now want to listen to Mongolian throat singing on your Spotify app based on your previous behaviour. Not impossible but less statistically accurate. The second difference relates to the cost of a mistake. Whilst the agencies use very much the same algorithms as those used by commercial entities, the impact of the decision by the algorithm is very different. The fact that I might get an advertisement for ice cream while I'm on a diet, or an irrelevant response from a chatbot when I really need to change something on my flight ticket, is annoying but considerably different from the decision to arrest me or even to kill me based on a mistake by an algorithm. Therefore, it is the combination of a higher risk of making a mistake due to the unique behaviour being predicted, together with a higher risk involved in the response to such a prediction that makes the use of AI by security agencies in predicting individual decisions unique. This research will also not focus on the general use of AI algorithms by security agencies as AI algorithms can be used by security agencies to help predict and discover patterns of various actions, from suspicious money transfer patterns that might suggest the funding of criminal activity, to analysing satellite images that can indicate an imminent attack by a militia military group. Even if some of the issues that appear in this research might also be relevant to the use of those algorithms, they do not attempt to predict unique individual human behaviour that leads to a set of very specific questions, both regarding our ability to predict a unique singular human behaviour like an individual attack and the actions we can take against an individual when the algorithm predicts an attack. This, therefore, is the very clear boundary of this research: the use of AI algorithms by security agencies to predict

individual unique behaviour such as a violent attack on others. As for the algorithms themselves, this research limits its scope to the ones active until 2023. The rise of powerful LLMs and their impact on counter-terrorism predictions are not discussed in this work. Having said that, I believe the discussion and conclusions are also very relevant to those algorithms.

As for the case study, the research focuses on the actions of the Israeli Security Agency (ISA) in creating and operating an AI algorithm to predict individual Palestinian attacks in the Occupied Palestinian Territories during the years 2015 to 2017. The methodological chapter will elaborate on the reasons for choosing this specific time frame and geographical area. In a nutshell, this research focuses on events that, according to the ISA, began in October 2015 with a series of individual attacks by young Palestinians against Israelis and the activation in January 2016 of an AI tool that was aimed to predict the next attacker. The period of the research covers the time before the wave of attacks started, the period when the attacks took place before the activation of the AI tool, the period when the AI tool was active and the violence continued, and the period the AI tool was active and the violence had ended.

Thesis Structure

This structure of this work is based around three main questions: The first question is, what are the conditions that might bring a counter-terrorism agency, and that brought the Israeli Security Agency (ISA), to develop a predictive AI to target individual attackers? The second question is, how did the ISA AI tool work and was it effective, in a narrow sense, in identifying potential attackers and, in a wider sense, in reducing the violence? The third question is, can there be a

better way to create and activate such a predictive tool, and should we even try to do it? The chapters of the research follow this logic.

Apart from this introduction, this research contains another seven chapters. Chapter Two presents a short and focused summary of the relevant literature regarding the use of predictive algorithm tools to identify potential individual attackers in counter-terrorism. The chapter starts by defining the current literature regarding the potential attackers that AI tools are looking for. It discusses the rise in popularity of the term 'radicalisation' as the new bucket term to ask questions about sources of terrorism, then reviews the literature regarding 'online radicalisation' and its connection to the literature on 'lone wolf terrorism'. After defining the literature regarding who the target is, the chapter discusses the predicting side and the challenges involved. The discussion starts with reviewing the challenges that arise from the literature regarding the abilities of humans to predict other human's behaviour in general, the possibilities of using statistical methods to predict human behaviour, and the possibility of using statistical methods to predict a rare phenomenon such as terrorism. Then, the discussion moves to review the attempts to use algorithms for predictive policing, focusing on the literature on 'person-based predictive policing' and its associated challenges. The chapter ends with a short review of the literature regarding the specific case study, the ISA tool that was used to predict individual Palestinian attackers. After defining the current literature, I move to discuss the theoretical framework of the research in Chapter Three. This chapter starts by presenting the general ontological and epistemological concepts of this research. It discusses the issues arising from my positionality as a former human rights defence lawyer who challenged ISA agents on a regular basis in the Israeli military courts and how those have brought me to adopt an ontological

position and theoretical framework from theories in critical terrorism studies. The chapter then briefly reviews the different theoretical advances made by critical terrorism studies and then focuses on critical counter-terrorism theories. The chapter then moves to theories regarding division of moral labour as a prism to explore questions of moral agency when collaboration between humans and algorithms have a clear moral impact. The discussion then narrows to presenting Sondre Lindahl's model of counter-terrorism as presented in his book *A Critical Theory of Counter-Terrorism* (Lindahl 2019).

After presenting the theoretical framework, the research defines how these theories fit into the research methodology in Chapter Four. The chapter starts with presenting the methodological structure of the research and my decision to first examine the case study through the eyes of the Israeli Security Agency and only after that from a critical point of view. The second part of the chapter describes how the data was collected, focusing more on the challenges that arose following the freedom of information request that reached the Israeli Supreme Court, as well as in obtaining the interviews with former ISA agents. The last part of the chapter discusses more broadly the methodology of the research, the data analysis techniques that were used to obtain as many observations as possible on the available data with a particular focus on the AI-generated indictments against Palestinians, and a discussion about the choice of Sondre Lindahl's critical terrorism studies model of counter-terrorism as a way to assess a theoretical predictive AI tool that could predict individual attacks. Those four chapters that set up the research basis lead to the data and discussion itself. In Chapter Five**,** which is the first case study chapter of the research, I discuss the Israeli security agencies' narrative around the creation and activation of the preventive AI tool. Firstly, the chapter discusses the scope of the case study in terms of time,

geography, and population, detailing the Israeli security services' definitions of what happened and how that has led to the creation of the ISA's AI tool. Later, the chapter presents a short background history of the ISA's operations in the Occupied Palestinian Territories and its abilities and activities prior to what they identify as the beginning of the wave of individual violence in October 2015. Next, the chapter presents the ISA's analysis of the reasons that led to the beginning of the wave using the ISA's own reports and articles. This is followed by an analysis of the challenges all Israeli security forces encountered in dealing with individual attacks, when compared to organisational-based violence and the different methods that were used to try and stop it. The last part presents the decision to create the AI tool, the efforts to collect the relevant data to train it, the parallel efforts that were carried out by other security bodies, and the activation and operation of the tool. Throughout this analysis, the chapter presents the statistical data collected by the ISA regarding what are identified by them as individual attacks and their patterns. This chapter discusses the ISA story of the AI preventive tool. However, the story, as it is told by the ISA, does not reveal the signals the AI tool was searching for and the end result of the tool identification. In order to try and examine those, in Chapter Six I discuss and examine, from two aspects, the criminal indictments against Palestinians that were based on the ISA's AI tool. The first aspect is to better understand the signals the AI tool looked for to reach a conclusion on a person, and the second discusses the legal system's response to the AI-based indictments. The chapter starts with a short historical background on the Israeli military legal system in the Occupied Palestinian Territories and describes the military courts and the military prosecution that oversaw the writing of those indictments. Later, the chapter focuses on the ways in which this research identifies those indictments, as they do not directly state that they

were created based on the information received by the ISA's AI tool. Then, the chapter discusses the observations that can be gained by a textual analysis of the AI-based indictments, and what these say about the algorithm's definitions and outputs. The last part of the chapter discusses the Israeli military courts' response to those kinds of indictments and the Israeli military prosecution's response to the courts' decisions and the impact those decisions had on the people that were identified by the ISA AI tool and arrested and indicted.

After presenting the case study and all the data collected in relation to it, Chapter Seven discusses the findings from the previous two chapters and uses them to critically assess the ISA's definitions and assumptions regarding its AI tool. The chapter begins by questioning the ISA's assumptions about waves of violence, as described from its perspective in Chapter Five. It addresses issues such as the causes of violence outbreaks and the criteria used to define an "attack" when constructing the initial dataset for the AI tool.

In the second part, I examine why the development of such a tool was so appealing to the ISA and other security bodies, focusing on the theory of the division of moral labour. The third part evaluates the effectiveness of the ISA's AI tool—both in the short term, in identifying potential attackers, and in the long term, in reducing overall levels of violence.

In Chapter Eight, I use the Israeli case study alongside other examples to envision the optimal conditions under which a theoretical counter-terrorism predictive AI tool could more accurately identify potential attackers from a statistical perspective, while remaining legally compliant within liberal democratic frameworks. I then explore the theoretical deployment of such mechanisms through Lindahl's model of counter-terrorism. The tool is analysed according to the model's five components—key assumptions, basic principles, strategies and tactics, prevention,

and evaluation—to assess whether, under optimal conditions, it can be aligned with critical counter-terrorism thinking.

Finally, the Conclusion chapter summarizes the research findings and proposes a possible ethical framework for the use of predictive AI technologies to counter individual acts of violence. It also discusses the broader theoretical implications and future possibilities of such technologies.

The Findings of this Research

To conclude, this research has reached several findings regarding the questions asked throughout it. Firstly, this research finds that the race by counter-terrorism agencies to create a predictive AI tool that would identify an individual attacker was fuelled by a combination of things, including the frustration of those agencies with the constant failures in preventing those attacks using classic counter-terrorism methods, the hype around artificial intelligence technology, and the need to remove some of the moral responsibility in attempting to pre-identify an attacker. Secondly this research shows that although operating in almost optimal conditions for predictive AI, there is no evidence that the ISA's AI tool was effective in identifying individual attackers, or in stopping the phenomenon of individual attackers, or in reducing the level of violence in the region. Thirdly, this research questions the ability of such AI tools, even if operated under future conditions that will answer some of the existing technical and legal hurdles, to be effective in reducing violence when operated in democratic settings. The research concludes with a suggestion. Instead of trying to identify individuals, the growing powers of artificial intelligence could be harnessed to try and pre-identify the regions where socio-political tensions could lead

to the rise of individual attacks, and to try and solve those tensions via socio-political means prior to the outbreak of violence.

Chapter 2 – Literature Review: From Defining Online Radicalisation to the Possibility of Algorithms Predicting Individual Attacks

The question of whether an AI algorithm could or should predict a potential individual attacker is asked in a particular world and conditions. No one is expecting an algorithm to issue a warning when an off-grid loner who lives in a wooden cabin in the forest decides one day to take his hatchet and kill people based on a manifesto he wrote to himself in his notebook with his pencil. The challenge for the algorithm is very clear: could an algorithm identify out of the millions of people that interact with online extremist and terrorist content, the ones who have been both radicalised enough and who carry all the signals that they are ready to act based on their radicalised beliefs?

This chapter will review the relevant literature regarding this specific question, starting with the research on radicalisation, online radicalisation, and online radicalised individual attackers, moving to the research on the ability to predict human behaviour and specifically individual human behaviour, and finishing with the research regarding the ability of algorithms to predict individual human behaviour, in the policing and counter-terrorism sphere.

The Concepts of Radicalisation, Online Radicalisation and Individual Attackers

Radicalisation and counter-radicalisation have become extremely popular terms for researchers and policymakers, especially since the 2004 Madrid and 2005 London bombings. According to M.S. Elshimi, 'radicalisation did not exist before 2004' (Elshimi 2017, 21). The term was created,

as Peter Neuman identified, post the 9/11 attacks to discuss the root causes of terrorism in a way that will not justify terrorists:

> 'Following the attacks against the United States on 11 September 2001, it suddenly became very difficult to talk about the 'roots of terrorism', which some commentators claimed was an effort to excuse and justify the killing of innocent civilians. Even so, it seemed obvious (then) that some discussion about the underlying factors that had given rise to this seemingly new phenomenon was urgent and necessary, and so experts and officials started referring to the idea of 'radicalisation' whenever they wanted to talk about 'what's going on before a bomb goes off'. In a highly charged atmosphere following the September 11 attacks, it was through the notion of radicalisation that a discussion about political, economic, social and psychological forces that underpin terrorism and political violence became possible again' (Neumann 2009, 4)

The 2004 and 2005 terrorist attacks in Madrid and London have led the EU to adopt the growing radicalisation discourse into its official policies; for example, the Commission's directorate general published, in 2005, a communication named: 'Terrorist recruitment: addressing the factors contributing to violent radicalisation' where it defined "Violent radicalization" as a 'phenomenon of people embracing opinions, views and ideas which could lead to acts of terrorism' (DG 2005, 1). The 2004/5 attacks also led to the radicalisation discourse, at least initially, being adopted by governments, agencies and scholars as a clean way to speak about the methods of dealing with home-grown Islamist political violence (Kundnani 2012). Very much like

the research on terrorism, the research on radicalisation lacks agreement on a central definition for 'what is radicalisation?' A report by experts for the Australian Government summarised it nicely: 'About the only thing that radicalisation experts agree on is that radicalisation is a process. Beyond that, there is considerable variation as to make existing research incomparable' (Nasser-Eddine and Caluya 2011, 13). This variation regarding the question of what the process of radicalisation is, can be found in the different models of radicalisation found in research. Moghaddam, in 2005, came up with a psychological model of the process that makes a person act. According to him, the process is like a staircase, each staircase out of the six described leans on the previous one and only those that have passed through all of them and reach the last staircase act violently (Moghaddam 2005). The Moghaddam model, although criticised for being too linear and lacking empirical data to justify (Lygre et al. 2011), is a basis for discussing a model that will define the process of radicalisation. One of the things that was lacking from the Moghaddam model, according to some researchers, was the fact that the process of radicalisation is not a totally individual one and it needs some conditions to be in place at a group level, in order to support the individual process. This led McCauley and Moskalenko to develop a 12-stage process which also includes the required stages at the group and mass level, such as the development of group hatred and competition, to allow the individual steps to continue towards radicalisation (McCauley and Moskalenko 2008). The continuous data regarding the personal agendas of individual attackers has exposed, that there was a need to tie in the existing models to specific personal characteristics and circumstances, as too many people who fitted the models did not choose violence at the end of it. A more sophisticated model, called the 'Significance Model', that tried to describe a non-linear process that includes personal motives, in parallel to

ideological and social processes, was developed (Kruglanski et al. 2014). However, its complexity made it very difficult to confirm or disprove. Academic attempts to create a model for the radicalisation process therefore moves between two scopes. On one side, a clear linear process which is questionable in its accuracy and, on the other, a complex multi-levelled model which is difficult to verify. A similar issue can be found regarding the academic definition of radicalisation. Out of the many interesting attempts to define radicalisation, one admirable attempt was made by Schmid, with a view to including most of the definitions relating to radicalisation:

> 'An individual or collective (group) process whereby, usually in a situation of political polarisation, normal practices of dialogue, compromise and tolerance between political actors and groups with diverging interests are abandoned by one or both sides in a conflict dyad in favour of a growing commitment to engage in confrontational tactics of conflict-waging. These can include either (i) the use of (nonviolent) pressure and coercion, (ii) various forms of political violence other than terrorism or (iii) acts of violent extremism in the form of terrorism and war crimes. The process is, on the side of rebel factions, generally accompanied by an ideological socialization away from mainstream or status quo-oriented positions towards more radical or extremist positions involving a dichotomous world view and the acceptance of an alternative focal point of political mobilization outside the dominant political order as the existing system is no longer recognized as appropriate or legitimate.' (Schmid 2013).

Of course, this broad and interesting definition has some weaknesses. The most obvious one is that, in an attempt to capture all of the possibilities around radicalisation, some political non-violent resistance behaviours might be included. In direct contrast to this definition is the Israeli Security Agency (ISA) definition of radicalisation, as given to me in an interview with a former ISA agent: 'the process that a non-involved Palestinian is going through when he chooses to act violently against Jews'.[2]

Like any other social phenomenon, extremism and terrorism have also found their way to the online sphere and especially social media platforms. As Conway already identified in 2016: 'the question is no longer if the Internet has a role to play in contemporary violent extremism and terrorism, but the more pertinent issue is determining its level of significance in contemporary violent radicalisation processes' (Conway 2016, 81). Following that statement, it is not surprising that the amount of research on the online activities of violent actors has grown immensely (Brown and Pearson 2018, 149). At the start, most of the efforts were directed towards the actions of radical Islamist groups, but as right-wing terrorism began to be acknowledged, mainly after the Christchurch attacks, significant academic research has also been dedicated to those (Blackbourn et al. 2019).

Online radicalisation has been identified as having a few significant differences compared to classical face-to-face radicalisation. For example, online anonymity presents less of an initial risk of users starting to interact with radicalised content, the younger population are more easily targeted, hateful views are very accessible, there are simple ways to disconnect from such

[2] From an interview with retired ISA agent 'BC' on 28/04/2022.

interactions, and there is no clear organisational hierarchy and fewer obvious incentives to evolve from online activity to action (Sageman 2008).

Several factors can explain the success of online radicalisation; the ability to reach incredible amounts of people in a vast geographical space, the creation of 'echo chambers' which normalise and advance extreme views, the exposure to violent actions carried out by radicalised activists that end in their death and the resulting support and praise they get in the 'echo chambers' that have a seducing effect for those seeking recognition and a normalising effect on the concept of death during an action; the exposure to graphic violent videos of war atrocities that can create moral outrage; and the ability of the users to reinvent themselves as more extreme characters using the anonymity of the web (Neumann 2013; Bastug et al. 2020). As social media has evolved and many actors have joined an arena that Facebook once ruled, online radicalisation has changed, too. As Conway has identified, it is clear that the radicalisation process on Facebook is different from that on Twitter or Telegram, and, therefore, each of them might work differently (Conway 2017).

The advancement of online radicalisation theories which brought Sageman to claim that 'face-to-face radicalisation has been replaced by online radicalisation' (Sageman 2009, 41) has led to a few researchers questioning the dichotomy between online and offline radicalisation, showing both empirically and theoretically that there is a constant connection between the two (Gill et al. 2017; Whittaker and Charmin 2021; Hamid and Ariza 2022).

Whittaker suggests that 'the question should not be "do terrorists radicalise online?"' but instead 'what role do information environments play in radicalisation?' This reframing forgoes an

online/offline dichotomy, which is neither empirically nor ontologically defensible (Whittaker 2022, 34).

From Online Radicalisation to Individual Attacks

The connection between online radicalisation and individual attackers has been well established. Paul Gill, who conducted several empirical studies himself and with others regarding individual attackers, found that, from 2012 on, at least 76% of individual attackers used the internet to learn about terrorist activity, almost half of them proactively downloaded terrorist-related materials to their own devices, 29% communicated about terror activities with others, and 15% actively disseminated terrorist propaganda online (Gill et al. 2017). Gill and others also found that: 'There is little evidence to suggest that the Internet was the sole explanation prompting actors to decide to engage in a violent act. Instead, it was just one factor among many that helped crystallise motivation, intent, and capability at the same time and place. Our results further suggest that many went online, not to have their beliefs changed but rather to have them reinforced' (Gill et al. 2017, 114). The research also did not present any data on how many people have consumed similar content and have not chosen violence as their way.

A continuous effort has been made by several researchers to identify the specific traits of individual actors and to identify distinctive pre-attack behaviours (Spaaij 2010; Boyle 2013; Gruenewald et al. 2013; Gill et al. 2014; Ganor 2021). An attempt to consolidate all of the findings into a clear data-based recommendation was carried out by Schuurman and others (Schuurman et al. 2018). This resulted in four findings which, according to them, could help in the detection and prevention of lone attacks:

- ‘Lone actors tend to be poor at, or unconcerned with, operational security;
- They engage in leakage behaviour that allows others to glimpse their convictions and violent intentions;
- The majority of lone actors do indeed maintain social ties that are crucial to the development of their motivation and capability to commit acts of terrorist violence;
- Temporal analysis indicates that most of the elements that are crucial to the planning and preparation of a lone actor terrorist attack begin months, if not years, beforehand, which suggests that law enforcement and security agencies need not necessarily rely on last-minute indicators of an impending strike but, given sufficient data and a correct analysis of contextual specifics, can engage in the early detection, interruption, and prevention of lone actor violence’ (Schuurman et al. 2018, 1998).

Although these attempts to find consolidating factors regarding individual attacks, and especially attackers, have been criticised by critical thinkers for the dangers they bring with them (Ojanen 2010; Monaghan and Molnar 2016), these types of analyses likely fed into the idea of developing AI tools based on such findings that would help pre-identify such attacks prior to their execution.

Predicting Human Behaviour, Predictive Policing and Predictive Counter-Terrorism AI Tools

The idea behind the AI tool that is the basis of this research is its ability to predict unique human behaviour, such as an individual's decision to attack others before it happens.

What is artificial intelligence, and how could it help humans to predict behaviour, is a question with no clear agreed answer (Nilsson 2009; Giovagnoli 2013; Müller 2018). A comprehensive analysis of the challenges in defining AI was carried out by Wang (2019). According to him, the problem is defining the 'intelligence' part of artificial intelligence and how it can be recognised. In the article, he examines his own definition, which is based on the idea that intelligence is the ability to arrive at a result in an environment of insufficient knowledge and resources, and he compares it to the different perspectives of intelligence he identifies in existing literature (structure, behaviour, capability). He reaches the conclusion that he has also failed. In his conclusion, Wang emphasises:

> 'According to this analysis, there is no correct working definition of AI, as each of them has theoretical and practical values, so they are not wrong. However, all working definitions are not equally good when judged according to the criteria introduced at the beginning of this article. Though there is no such a thing as a perfect working definition, and I do not expect a consensus to form soon on which one is the best, at least the ultimate incompatibility among the perspectives should be recognized. It is still up to each researcher to choose how to use the term "AI" though it should be clarified when the result is discussed, with its implications understood well' (Wang 2019, 29).

Predictive AI, the topic of this research, is a good example of the problems in defining AI. The term 'AI' throughout most of this research will be used to describe an algorithm which has the goal of predicting a single human behaviour based on the former behaviours of that individual. Therefore, in this research, an AI tool will be considered as such if its operators believe it can

predict human behaviour. The question regarding its effectiveness in predicting, and the meaning of such an ability will be discussed, regardless of its definition.

The challenges in Predicting Human Behaviour

Before discussing artificial intelligence and its abilities, it is important to remember that we, as humans, are pretty bad at predicting such behaviour. Daniel Kahneman, dedicates a large part of his book *Thinking Fast and Slow* (2012) to presenting all the reasons why we, as humans, fail time after time to predict human behaviour. Our biases, overreliance on intuition, overconfidence, illusion of validity, and the general uncertainty around human behaviour which is influenced by numerous factors, all prevent us from coming up with an accurate prediction (Kahneman 2012, 199–255). Kahneman believes the answer for such a prediction could be with future artificial intelligence, if given enough unbiased data.

The question of whether statistical data analyses should be used to predict individual human behaviour has been discussed in the field of psychology since Paul Meehl published, in 1954, a book suggesting such statistical observation can predict some kinds of human behaviour (Meehl 1954). Meehl suggested that the statistical prediction of human behaviour better identifies long-term trends and failings when an environment suddenly changes. As an example, he mentions a situation where, based on statistical data, at a faculty of a specific university, the statistical analysis will predict who will be the first professor to go to a new movie that arrives on the screens; the statistics show that a particular professor is likely to do so; however, the statistical analysis is unaware that the professor has just broken his leg and therefore cannot go to the movies. These situations, later named 'broken leg' variables (Salzinger 2005), have been for many

years the psychologist's response to any claim that statistical data could be used to predict a singular person's behaviour. However, even psychologists cannot ignore the information revolution that the internet and its uses have brought with it; in a way, the amount of personal data each individual is sharing has flipped the 'broken leg' variables argument. As Puyvelde and others have identified, today, it is more likely that the information regarding the professor's broken leg would be found easier and faster by an algorithm than by a human (Van Puyvelde et al. 2017). The ability to analyse the big data that is gathered during our constant interaction with the digital world and make predictions regarding our future actions based on it, is the basis of online marketing (Alfiqra and Khasanah 2020). Today, it is not only that the algorithm will know that the broken-legged professor has broken their leg and cannot go to the cinema, but it will also offer him the right painkillers for his leg (Guay and Parent 2018). Having said all that, there is still a significant statistical difference between a broken leg event and terrorism. In the US alone, in the year 2023, around 6 million people were reported by hospitals to have broken a limb (u.osu 2025); that year, there were only 23 terrorism-related attacks in the US (IEP 2025). The uniqueness of the phenomenon of terrorism and its rarity has led to psychologists like John Monahan concluding that 'existing research has largely failed to find valid non-trivial statistically significant risk factors for terrorism. Without the identification of valid risk factors, the individual risk assessment of terrorism is impossible' (Monahan 2012, 19) meaning there is too little data and too many different variables to establish a statistical model of a potential terrorist.

The Race to Use AI in Predictive Policing

As always, clear academic statements that something is impossible have never deterred engineers and policymakers from trying. Attempts have been made to use computers to use

statistical data to predict crime since 1990. Bilel Benbouzid described the first attempts US police forces and private companies made in developing such systems, from the Philadelphia police force, which introduced the Philadelphia Crime and Mapping Systems (PHiCAMS) and the Crime Spike in 1997, which collected all of the data from all of the various districts and presented it visually in order to map potential crime areas, up to the popular 'PredPol', software which appeared in 2012 and was the first to suggest that it can predict where and when a crime will happen and can send a police patrol there to do a stop and frisk operation (Benbouzid 2019, 3–5). The growing popularity of predictive policing algorithms such as 'PredPol', 'Hunchlab', and 'Palantir' and their adaptation by police forces around the world have started to attract researchers' attention. Andrew Ferguson (2012; 2017; 2017) was one of the first to identify the different kinds of predictive policing to be quickly developed and implemented, and he discussed their possible impacts. Ferguson divided the development of predictive policing into three periods. Predictive policing 1.0 focuses on identifying the places and times of property crimes, like car thefts and burglary, with a view to increasing the police presence there. Those were chosen because they were common, very often reported, and had a potentially big and positive impact on the non-criminal population. Those early programmes, which mostly concluded in producing specific daily patrol routes to police offices, have shown early results of success across the US (25% drop in burglary in LA). It was not a wonder that police and tech companies started to think they could do better (Ferguson 2017, 2016). Predictive policing 2.0, according to Ferguson, tried to focus on predicting the where and when of violent crimes such as aggravated assaults and gun violence. Those algorithms offered a risk model according to the crime; for example, the model identified that the following areas are more likely to see gun violence:

'locations of drug arrests, proximity to 'at-risk' housing developments, 'risky facilities,' locations of gang activity, known home addresses of parolees previously incarcerated for violent crimes and/or violations of drug distribution laws, locations of past shooting incidents, and locations of past gun robberies' (Kennedy et al. 2011, 345). The focus of policing 2.0 on violent crimes led to the algorithms narrowing police activity to where those crimes are most likely to happen, usually the poorest neighbourhoods, where gang and drug activity were most common. That focus on specific violent areas once again led to a drop in crime rates, and created a set of different questions on why police forces had to wait for an algorithm to act, as those areas were known as violent areas before any algorithm was created, and had the focus on poor areas led to over-policing of the innocent population there (Eubanks 2017). Predictive policing 3.0 jumped from predicting where and when crimes will happen to who will likely commit the crime. These algorithms, which are closely related to the case study and the research question, were developed to create, based on the growing personal data that can be collected thanks to an individual's digital footprint, a personal risk assessment that could suggest who has the potential to commit a crime. The excitement regarding those new methods reached the US Bureau of Justice Assistance, and led to a report about these 'smart new approaches', mentioning, for example, the LAPD's LASER (Los Angeles Strategic Extraction and Restoration) project that was developed to create lists of targeted individuals to be placed under surveillance because their profile as gang members and repeat offenders suggested it (Braga et al. 2014). In parallel to those developments, Palantir, then a new surveillance software actor competing with PredPol, launched a network analysis algorithm in the city of New Orleans in order to reduce gun violence. According to Palantir, their algorithm identified 1% of the city population as having the potential

to be involved in gun violence. According to the Palantir developers, the ability to focus on the right population led to a drop of 20% in the city's murder rate (Iliadis and Acker 2022). Those programmes, which have grown in popularity and ambition globally over the years, were the basis of what was later named 'person-based predictive policing,' and as Ferguson identified, there was a 'rapid evolution from place-based property crimes to place-based violent crimes and then to person-based crimes. This evolution has largely gone unchallenged, even though the social science justifications for the different crime types remain contested' (Ferguson 2017, 1115).

Reaction to the Rise of Person-Based Predictive Policing

What Ferguson identified as unchallenged has changed dramatically over the following years; the increased level of discourse around AI capabilities and big data algorithms has not skipped predictive policing. More and more critical voices have begun to emerge, questioning both the success stories presented by the police and software companies and the social impacts of those tools. Various NGOs started writing reports criticising the different practices being developing globally. A Human Rights Watch report from 2017 analyses the Chinese police software that gathers, among other data:

> 'patient records - including names and illnesses - obtained from the National Health and Family Planning Commission; names and causes of petitioners - individuals who complain to the government, usually for official abuses - from the State Bureau of Letter and Visits; and the names and addresses of individuals convicted of crimes from the Bureau of Justice. The Police Cloud will also aggregate company data, including user names and their

> IP addresses from telecoms companies; usernames of their social media accounts (wechat, weibo, QQ, and email) from internet forums; and senders' and receivers' names, phone numbers, and declared package content from delivery companies…..navigation data on the internet, [and] the logistical, purchase and transaction records of major e-commerce companies' (Human Rights Watch 2017, 3).

This vast data is used, according to the report, to alert any unusual activity, discover relationships that were not apparent to the police, and surveilling specific groups of interest, from terrorists to people who tend to cause disturbance. The report concludes that those systems are breaching the right to privacy, might chill freedom of expression, and might be used to target specific populations (Human Rights Watch 2017, 7). An Amnesty UK report from 2018 criticised the London police 'Matrix' algorithm that provided individual risk assessments to categorise those they identified as gang members in the city according to how dangerous they were, where 'red nominals' are those the police consider most likely to commit a violent offence; 'green nominals' pose the least risk' (Amnesty International UK 2018, 2). The report finds that, although 72% of the youths identified as 'red nominals' were black, black youths are only 27% of those youths responsible for serious violence in the city. The report continues by showing the impact of such labelling on young people and concludes that:

> 'Gangs Matrix is unfit for purpose: it puts rights at risk and seems not only ineffective but also counterproductive. Systems for gathering and sharing intelligence on individuals suspected of violent crime must be fair, implemented in accordance with human rights law, and have robust oversight mechanisms' (Amnesty International UK 2018, 4).

Hanna Couchman carried out another extensive report for the UK Liberty Organisation, surveying all the different uses of algorithm predictive policing in the UK (Couchman 2019). By sending 90 freedom of information requests to all of the police forces in the UK, the report revealed that at least 14 of them were using some kind of predictive policing algorithm. The report examines the different UK programmes and reaches the conclusion that: 'predictive policing programs entrench pre-existing inequalities while being disguised as cost-effective innovations in a time of austerity and their use puts us at risk' (Couchman 2019, 4). All three civil society oreganisation reports identify the same four repeating issues that make up the common criticism of person-based predictive policing: a) The level of surveillance that is needed infringes the right to privacy; b) It is not effective and brings a high rate of false identification; c) It is based on biased data and therefore produces biased results; d) The algorithm results work as a 'black box' and therefore the recommendations cannot be assessed and regulated (Human Rights Watch 2017; Amnesty International UK 2018; Couchman 2019).

In parallel to the NGO reports, person-based predictive policing has begun to draw the attention of academics. Ferguson, who was leading the research and held a critical point of view on the use of person-based predictive policing in the US, identified five main 'potential systematic vulnerabilities' in predictive policing and person-based predictive policing specifically. The first vulnerability identified by Ferguson is caused by the potential problems in the data gathered by the algorithms in order to make the prediction. This kind of data could be lacking, mistreated, or handled by humans and, of course, biased as it is based on the previous biases of the police force (Ferguson 2017, 1147–50). The second vulnerability Ferguson identifies in predictive policing is a methodological one with a focus on the lack of ability to validate the results of the person-based

algorithm recommendations and differentiate between correlation and causation in the impact of the algorithm (Ferguson 2017, 1155–57). The third vulnerability Ferguson raises is regarding the lack of social science research and theory to back up the ability to statistically predict the criminal risk of an individual in a specific moment (Ferguson 2017, 1162–64). The fourth and fifth vulnerabilities are the lack of transparency and accountability, which is the result of working with an algorithm that does not explain how it reached the recommendation or risk rating it provided and cannot be held accountable for its mistakes (Ferguson 2017, 1166–70). Those five vulnerabilities, as identified by Ferguson, are routinely the basis of other follow-up research on person-based predictive policing. In the UK, Alexander Babuta acknowledges the possible benefits of incorporating algorithms to create statistical risk assessments as a labour-reducing tool but once again raises the vulnerabilities of biased data and lack of transparency as major hurdles. He identifies two types of transparency required when working with predictive algorithms: technical transparency, which makes it clear which data the algorithms use and what kind of processing is done to it in order to reach a prediction, and transparency of the process, which focuses on the process behind the action that is based on the prediction, once created (Babuta et al. 2018, 28–30). After researching the data from two algorithms used by the UK's Durham and West Midland police forces for individual risk assessment, Babuta (2019) identified the issue of bias as these algorithms' main problem in terms of efficacy and legal compliance. A more elaborate discussion on the legal challenges arising from the use of policing algorithms in the EU was carried out by Athina Sachoulidou (2023), who claims the growing use of these algorithms all over Europe is changing the base presumption of innocence of the individual. Similar claims are raised by Ales Završnik, who suggested a new prism to look at the use of AI

algorithms in the criminal law system, naming it ‘Algorithmic Justice’. His version of using the term was to create an umbrella discussion on ‘big data, algorithms and machine learning in the criminal justice domain’ (Završnik 2021, 629). However, the term was so successful that it later began to be used for any social justice/legal implication that the use of AI has on humans (Marjanovic, Cecez-Kecmanovic, and Vidgen 2021). Another aspect of person-based policing that has been researched, mainly in the US, has focused on trying to decode what the specific algorithms were programmed to do. As the algorithms were created by private companies for profit, they were naturally reluctant to share their code. The most researched algorithm was ‘Palantir Gotham’, which has been used by police forces both in the US and Europe. This algorithm, which provides wide services for preventive police needs, also has a person-based risk assessment aspect, which seems to be popular. However, academic attempts to validate its abilities vary too much in order to come to a clear conclusion on its benefits in comparison to other algorithms (Iliadis and Acker 2022; Gundhus and Wathne 2024; Wei et al. 2024). It seems that the basic vulnerabilities that were raised by Ferguson in 2017 about person-based police algorithms have been, in most of the cases, left unanswered even today, and although some regulation attempts have been advanced in some parts of the world (Levano 2024), the adaptation of person-based algorithms by police forces continues to grow all over the world.

Using Algorithms to Identify Potential Online Radicalised Individual Attackers

Although in many aspects, person-based police algorithms, which are used to create an assessment of a criminal, are very similar to assessing the risk of online radicalised individuals in order to identify potential individual attackers, they have some very distinct differences. First and foremost, the agencies that are dealing with it are usually specialised counter-terrorism agencies

outside of the police force, and operate in greater secrecy and, therefore, researchers have less access to data. Second, the relatively small number of attacks by radicalised individuals in comparison to violent crimes (23 terrorist attacks in the US in 2023 compared to 40k+ cases of gun violence (IEP 2025)) makes it, as already mentioned, very difficult to create a valid socio-psychological model that will fit potential attackers (Monahan 2012). Those differences can be seen in the literature regarding the use of AI algorithms by counter-terrorism agencies to predict individual attackers. The secrecy in which most counter-terrorism organisations are operating has led to a situation where there is almost no published empirical study to be found on predictive counter-terrorism AI tools designated to predict radicalised individuals, and most of the writing is focused on the theoretical questions that arise from using such tools.

Most of the existing research about using AI to predict a specific attacker is usually found in a more extensive discussion on the possibilities of using AI for national security. Babuta, who focused on predictive policing in the UK, has also examined the use of AI for national security needs (Babuta et al. 2020). The research, funded by GCHQ, examines how AI is used, or can be used, for national security in the UK and is based on open-source information and interviews with agents. When it comes to the question of what Babuta calls 'behavioural analytics', which is the ability of an algorithm to predict specific human behaviour in the context of terrorism, the discussion is very much theoretical, with no reference to a specific system. It is focused again on the same identified issues that were identified regarding person-based policing; bias, transparency and accountability, whilst focusing on the lack of sufficient quality data to train these algorithms (Babuta et al. 2020, 13–16). Interestingly, although Babuta is critical of person-based policing, he still believes it could be used with the proper protective measures (Babuta

2019). However, his position regarding the use of similar technology in counter-terrorism is much more straightforward:

> 'In sum, the evidence reviewed for this paper suggests that it is neither feasible nor desirable to attempt to develop AI systems to 'predict' human behaviour at the individual level – for instance, for counter-terrorism risk assessment purposes' (Babuta et al. 2020, 16).

A similar position is voiced by Maria Schröter in her research on the use of AI in countering violent extremism (Schröter 2020). Based on interviews with researchers, policymakers, and private sector experts, the research identifies the benefits of AI in detecting online extremist content and even countering it. Schröter also reviews the individual efforts to use AI for those purposes in Canada, France, Ghana, Japan, New Zealand, the UK and the US (Schröter 2020, 34–40). However, regarding the question of whether AI can predict which of those exposed to online extremism will decide to act violently upon it, her conclusion is quite straightforward:

> 'It may have become obvious that a general AI, a system with super-intelligence, is not an option to forecast online radicalisation of individuals for two reasons. The first is technical: given the current status of AI technology, algorithms need vast amounts of data to make useful predictions on the future. Luckily, radicalisation and terrorism do not occur often enough to produce enough data for a general AI forecasting the behaviour of individuals regarding online radicalisation. The rate of false positives and false negatives would be intolerable. Investment into human resources would be more beneficial. The second reason is around privacy: a system observing real-time online behaviour of

> individuals, storing the data and analysing it would not comply with privacy standards in liberal democracies. It could potentially lead to the mass surveillance of society' (Schröter 2020, 25).

Those two reasons, technical inability due to insufficient data and legal constraints, when added to the already identified issues of person-based policing (data, bias, accountability and transparency), are usually the tipping point that makes researchers reject the notion of AI as a tool to identify and predict a specific attacker. Similar analyses were carried out by Fernandez and Alani, focusing on technical and legal aspects while reviewing the current literature and arriving at the same conclusion (Fernandez and Alani 2021). At the same time, a report by the United Nations Office of Counter Terrorism (UNOCT), which reviewed the potential use of AI by counter-terrorism agencies in South Asia, also marked those two issues (the technical and the legal) as the relevant blocks:

> 'Given the unpredictability of human behaviour and the current state of technological development, the application of algorithms to predict behaviour at an individual level is likely to remain of minimal value. Additionally, human rights experts and civil society organizations have pointed towards several ethical concerns regarding potential entry points for discriminatory judgement and treatment. The large quantities of data concerning an individual required for the algorithm to accurately function further give rise to concerns about the possibility of unwarranted mass surveillance' (UNCCT 2021, 24).

Timme Munk focuses on the statistical problems arising from the initial small data sets of terrorist attacks:

> ‘Terror is a low-frequency event (Harcourt, 2007; Knibbs, 2014; Rosen, 1954), and every single event can be seen as unique (Schneier, 2015), which means that the risk of low base rate fallacy (Horgan, 2008) and over-generalization increases. Methodological generalization is impossible, as the amount of data is too small (Mackenzie, 2015), and the result will always be underfitting or overfitting (Horgan, 2008). Specifically, there is not enough data to build a model or to train the model to make a meaningful prediction (Kaufmann, 2010). The problem is that there is no clearly defined pattern or statistical possibility of defining what can and must be seen as attempts, and failed attempts are often kept secret as an essential characteristic of criminal activity (Horgan, 2008; Jonas and Harper, 2006; Knibbs, 2014; Rosen, 1954). Any algorithmic classification will result in false negatives as well as false positives (Ananny, 2016; Diakopoulos, 2015; Kraemer, et al., 2011; Silver, 2013) and, as described previously, the size of these groups depends on the relationship between the total population and the population sought’(Munk 2017, 7).

Following on from that, Munk reaches the conclusion that: ‘The use of predictive methods to predict terrorism is therefore ineffective, risky and inappropriate, with potentially 100,000 false positives for every real terrorist that the algorithm finds’ (T. B. Munk 2017, 10). Christopher Wall agrees with Munk regarding the current status of the algorithms in identifying terrorist attacks: ‘this inability to comprehend variation means that many algorithms are ill-suited for detecting the type of black swan events that characterize terrorism’ (Wall 2024, 7). However, Wall is willing to assume that all the technical aspects of such a potential tool could be solved in the foreseeable future and focuses on the legal reasons that will prevent the activation of such an algorithm.

According to Wall, the problem that will not be solved is the threat to privacy. For such an algorithm to be efficient in recognising a potential individual as an attacker in the US, for example, the algorithm will need to collect an immense amount of private data on all of the citizens, which will require a change to the US constitution (Wall 2024, 14).

A relatively less deterministic point of view regarding the ability to create a functioning and legal predictive AI system that could effectively identify a radicalised attacker can be found in a research paper by Kathleen McKendrick for Chatham House, the UK's Royal Institute of International Affairs (McKendrick 2019). While McKendrick acknowledges the four basic problems (i.e. data, bias, accountability, and transparency) of predictive AI, she challenges the fact that the two extra problems of counter-terrorism AI, the statistical one and the privacy one, are making it impossible to use. According to her, those problems are interlinked and allowing better access to data could also solve the statistical problem:

> 'Existing assumptions – such as the beliefs that broader access to data is always deleterious for human rights, and that centralization of analysis is inherently bad – should be reviewed in the light of technical possibilities for controlling powers of access and increased transparency. The development and use of predictive capabilities can be a valid justification for wider access to, and use of, public data, provided that models are thoroughly validated before use, that the initial stages of analysis are automated, and that technical measures of control are in place to prevent misuse. Continued access to any data for the purposes of countering terrorism should be contingent on the ability to

> derive sufficient predictive value from those specific data – meaning that proportionality of access is directly linked to fulfilment of a legitimate aim' (McKendrick 2019, 34).

McKendrick, a senior officer in the British army, may present a unique position in the academic realm. However, the continuous efforts to advance such technologies worldwide show that her position is not unique in the counter-terrorism agencies' sphere, as attempts to use such technologies for counter-terrorism needs still continue in many places such as, for example, in the US with Palantir technologies (Jensen et al. 2020), and in the EU with its Frontex program (N. Howard 2024) .

The Israeli Counter-Terrorism AI Algorithm

In 2019, the Israeli army's Dado Centre for Interdisciplinary Military Studies published Issue C of *Routine Security*, its academic journal. The issue was dedicated to 'the campaign between the wars' and was focused on how the Israeli security forces dealt with the phenomenon of individual Palestinian attacks against Israeli targets in what was called the 'knife intifada' or, by the army, 'Maarehet Godel Hashaa'. The issue included articles from several military and intelligence officers presenting their point of view on how and why this military campaign succeeded, in their point of view (Carmeli 2019; Numa and Liraz 2019; Goffman 2019). A unique article was written by Arik Barbing, an ISA senior agent and Or Glik, a military research scholar, describing the ISA approach to the events. It was unique as it was the first time the ISA was invited to publish in this kind of publication, but more importantly for this research, it was the first time the ISA officially acknowledged it used AI technology to predict who could be the next individual attacker (Barbing and Glick 2019). In the article, Barbing, who was heading the development of the ISA AI project,

describes the reasons that led to the creation of the AI tool, the profile that was created of the potential attacker, the data collected by the algorithm and the overall results of activating the tool. Barbing claims that by analysing the radicalising social media content young Palestinians were consuming and creating together with other signals collected by the algorithm, the AI tool could identify who has the highest risk of attacking and when. Ganor, an Israeli terrorism scholar, has analysed Barbing's article, and although acknowledging the challenges usually associated with predictive AI as described above, sees this latest development as nothing less than a 'revolution in counter-terrorism' (Ganor 2019, 605). Another reference to the ISA AI predictive tool can be found in a book published by Brigadier General Y. S.,[3] which deals with the potential of the co-operation between advanced AI technologies and human agents (Y. S. 2021). The ISA tool is presented as one of the great success stories, claiming that the AI's ability to collect data and predict the attacker is equivalent to using '20,000 human analysts for 200,000,000 years' (Y. S. 2021, 76).

What is Missing in the Literature?

There is plenty of research about the radicalisation power of online content, and although academics are split regarding the question of the role of this content in the decision to act violently, there is a general agreement that it is not insignificant. There is also a growing body of work regarding person-based predictive algorithms targeting crime using varied signals, even though significant concerns were raised regarding the safeguards needed for it to be effective,

[3] Y. S. was identified in 2024 by *The Guardian* as Yossi Sariel, the commander of the Israeli intelligence technological unit 8200. It was the unit under his command that was mainly blamed for the lack of warning signs for the surprise attack by 3,000 Hamas militants on 7 October 2023. (https://www.theguardian.com/world/2024/apr/05/top-israeli-spy-chief-exposes-his-true-identity-in-online-security-lapse)

legal and just. The few researchers that have examined the theoretical possibility of AI predictive algorithms identifying who among those radicalised will be likely to individually act violently have concluded that, beyond the usual issues regarding predicting human behaviour, predicting violence by a radicalised individual is impossible statistically and legally. However, counter-terrorism agencies have been trying to develop such an algorithm, and there has been no attempt until now to evaluate their success or failure academically. The Israeli case study is unique because it has optimal conditions concerning the two unique problems of terrorism prediction. As will be detailed later, the case study is focused on a specific and relatively small population; the number of similar attacks carrying similar signals was, relative to terror cases, significant, and the ISA do not see itself limited in collecting any available data on the Palestinian population. Empirically researching a predictive AI tool that operated under those conditions could evidence – or, indeed, disprove – many of the assumptions raised in the research so far.

Chapter 3 - Positionality, Ontology, Epistemology and Relevant Theoretical Framework

In this chapter, I present the general ontological and epistemological foundations of the research, along with the theoretical framework that underpins it. I also address issues arising from my positionality as a legal scholar and former human rights defence lawyer, who regularly challenged ISA agents in Israeli military courts. This discussion leads into the second part of the chapter, which explores questions emerging from the use of AI in decisions with clear moral implications—such as employing predictive AI algorithms to determine whether a person should be arrested. It begins with a discussion of the general concept of moral agency and the theory of the division of moral labour, and then narrows its focus to situations in which moral responsibility is shared between humans and algorithms, examining the resulting impact on human moral agency. The third part of the chapter considers how counter-terrorism efforts can be assessed, focusing specifically on Lindahl's counter-terrorism model, which will be used throughout the research to evaluate the capabilities of predictive AI tools in counter-terrorism contexts.

CTS and My Research: Observations on Reflexive Positionality

There is general agreement that critical researchers must first challenge their previous assumptions and reflect on the position they are coming from concerning their research subject. The subjects of ‘terrorism’ and ‘counter-terrorism’ usually interact with compelling emotional and political positions as they combine, in most cases, violence, conflict and ideologies, and therefore, it is rare for a researcher to explore these subjects without some preconceptions. My personal experiences have certainly impacted on my research position. First and foremost, I need to acknowledge that I have an apparent political agenda about the Israeli/Palestinian conflict that

supports the right of Palestinians to live freely in a country that respects all of their rights as a free nation, rights that the Israeli occupation has oppressed for decades. This political agenda intensified during my years as a lawyer in Israel/Palestine. As described in the Introduction, before the beginning of the research, I was a defence attorney, mainly representing Palestinians who had been brought to the Israeli military courts and suspected of committing security offences. Although I represented hundreds of clients over the years, and although some of them were suspected of serious security offences, I never felt that I had met a terrorist. I met people who committed horrible, violent crimes, I met devoted political and religious ideologists, and I met good and evil people. Still, there is something in the intimacy created between lawyers and their clients that always made me see a whole person in front of me and not a symbol or a representation of a specific position. Each person I represented had specific and unique circumstances that brought them to the courts. Many of them were brought through no fault of their own and as part of the general Israeli attempt to control and subdue the Palestinian population. Some of them were arrested, although their actions were not violent and were only carried out as part of their need to be heard. Some of them, a small group of people, chose violence. Even in that small group, each person had a particular set of circumstances that led to them choosing violence. Some were rich, some poor, some highly educated, some illiterate, some were leftists, some right-wing, some religious, some secular.

It was no surprise then that when the first AI cases came to court, I was highly sceptical of the ability of AI to create a model based on people who are about to commit a violent act, and not only because of the technological element of it. My personal experiences have led me to believe that such a model cannot be effective, and I found the idea that such a thing is possible - based

solely on a person's digital activity - ridiculous. This view about my clients was also reinforced by my academic legal background, working mainly in the Israeli military courts, which represent such a clear example of the court as a power structure; I was naturally drawn to theories of Critical Legal Studies (CLS) and my masters examined the Israeli military court's jurisprudence regarding international law, and the ways the courts restructured its interpretation of it in accordance with the Israeli army's needs (Ramati 2019). Naturally, when I searched for a theoretical framework for the research, I was drawn to Critical Terrorism Studies as these scholars were questioning the same issues I was questioning regarding the understanding of terrorism and counter-terrorism.

Terrorism Studies, Critical Terrorism Studies and Critical Counter-Terrorism Studies

Terrorism Studies have developed as an academic discipline together with the formulation of modern terrorism as a concept (Stampnitzky 2016). There are a couple of overviews that explore the history of terrorism studies, such as those by Silke (2003) and Schmid (2011), and while they all mention that terrorism, as a phenomenon, has been researched by several individuals from the early 1970s onwards, there is a general consensus that 'it was only after the 9/11 attacks in 2001 that terrorism moved from the fringes of scientific interest to a subject of major attention' (Silke 2019, 1). However, those 'fringes', as Silke calls them, carried with them the main challenge that would continue to bother Critical Terrorism Studies' scholars years later, which is the tendency to focus on the state-centric, problem-solving approach (Jarvis 2009). Although this approach was used by the majority of terrorism researchers prior to the 9/11 attacks, some examples of critical research activity existed, which asked questions about sources of terrorism,

its definitions, and the motives and ideologies that guide terrorism researchers (Chomsky 1979; Herman and O'Sullivan 1989). This situation of having mainstream terrorism researchers who are state-centric and problem-solving oriented, along with a relatively small group of critical researchers who challenge them, intensified dramatically after the 9/11 attacks in the US. These attacks took place in a world that had developed massively in terms of access to live media, capturing and shocking the whole world. One of the many striking effects of that attack was the subsequent demand and suddenly available funding for terrorism research. As Silke has identified, this massive jump in interest in, and demand for, terrorism-related materials can be clearly seen in the number of publications relating to terrorism. For example, in the year 2000, there were 50 non-fiction books published with the word terrorism in the title, compared to more than 500 in the year 2002 (Silke 2009).

As already mentioned, even before the expansion of terrorism research after 9/11, there were critical voices in terrorism studies (Chomsky 1979; Herman and O'Sullivan 1989; Zulaika and Douglass 1996). However, the formation of Critical Terrorism Studies (CTS) as an academic field can be pinpointed to 2006 with the first symposium of critical terrorism scholars and the publication of 'The Core Commitments of Critical Terrorism Studies' (Jackson 2007) which summarised the symposium. This group of scholars, which has become more significant and more diverse, later published several collective works, such as 2009's *Critical Terrorism Studies: A New Research Agenda* (Jackson, Breen Smyth, et al. 2009), *Terrorism: A Critical Introduction* in 2011 (Jackson et al. 2011), 2016's *Routledge Handbook of Critical Terrorism Studies* (Jackson 2016), along with the founding of the quarterly journal *Critical Studies on Terrorism*. This body of literature has been growing as new voices and new ideas have joined, surveying a wide range of

topics from epistemological (Fitzgerald 2016) and ontological (Zulaika 2016) questions about terrorism through a historical analysis of terrorism studies (Stampnitzky 2016), to focus on specific kinds of terrorism (Lindekilde et al. 2019) and the counter-terrorism phenomenon (Lindahl 2019).

Jarvis (Jarvis 2016) has identified four significant criticisms CTS has of classical terrorism studies. When it comes to classical terrorism studies, he mentions that a) terrorism studies refuse to acknowledge terrorism as a social construction which leaves terrorism studies with a very narrow version of terrorism whilst ignoring other types of possible terrorism, like state terrorism; b) terrorism studies lack serious empirical data for various reasons; c) terrorism scholars have an unhealthy connection with the bodies (usually governmental) that sponsor them, which leads to the fourth and maybe most important criticism; d) terrorism studies are too focused on policy and problem-solving research, which once again aligns it with the counter-terrorism agencies' agenda (Jarvis 2016, 29-30).

When examining my pre-positions in accordance with CTS ontological positions, it seems that, even before I had read any CTS writing, I felt very comfortable with understanding all of those claims, as described by Jarvis. As I delved into CTS theories, I found that the position that some CTS scholars have adopted, the 'minimal constructivism approach' (Fitzgerald 2016), was a good fit for this research. The ontological view of this approach is that 'terrorism', 'terrorist', 'counter-terrorism', etc., are all social constructs that are time-, area-, and culture-dependent (Zulaika 2016). This is a simple argument to substantiate when looking at the historical development of conflicts associated with the term 'terrorism'. What was once considered 'terrorism' can be viewed later, in the same place, as a heroic war for independence. An act that could be

considered in one place as a terrorist act can be considered in another as a civil protest. A person can be defined at a particular stage as a terrorist and later win a Nobel Peace Prize for the same actions. What makes the 'minimal constructivism approach' interesting and relevant to the research is that it combines critical thinking of the situation with the ontological acceptance that political violence exists and needs to be addressed (Jackson et al. 2011). The address should be on the normative level that, as Martini says, should be: 'looking for and formulating new, less violent and more ethical and humane ways of dealing with and understanding political violence' (Martini 2021, 1). This critical and normative approach can also be found when examining CTS writings on counter-terrorism (Pettinger 2021) and led to Lindahl coming up with a model for counter-terrorism techniques (Lindahl 2019).

Based on those mainstream theories of CTS, this research also adopts the minimal constructivist approach as its ontological position. The research recognises that many terms at the heart of this research are, in fact, social constructs that are very unstable and impacted easily by time, circumstances and positionalities. Terms like security, terrorism, terrorists, radicalisation, counter-terrorism, counter-radicalisation, ideologies, justice and peace are exceptionally unstable when examined through the prism of an active conflict such as the Israeli/Palestinian one. For example, these terms can define very different things at the same time and in the same space, only based on the national identification of the person using them. Security, for instance, means something very different to an Israeli and a Palestinian. Having said that, the fact that those terms are unstable social constructs does not make them irrelevant to academic research, and that is especially true when it comes to their tremendous impact on much more stable terms in reality, such as death, violence, weapons, laws, prisons, computers, words etc. Societies use

those unstable terms to justify sending thousands of their members across the world to fight other societies, to kill and destroy properties, and to imprison and torture; therefore, understanding them is of vital importance.

The approach taken herein is to research the impact of those constructs on the more stable terms that the minimal constructivist approach recognises as being things in the real world that can be measured. For example, the research could recognise that the Israeli construct of the term 'security' means that no physical harm or property damage is done to Jews by non-Jews.[4] It could then check whether, under this construct, the actions of the Israeli security forces lead to more security, as defined by them, by examining the number of attacks on the persons and property of those who define themselves as Jews by those whom they define as non-Jews. This approach, which recognises the social constructs of terms like 'terrorism' and 'counter-terrorism' but also recognises the effects of those terms on stable social constructs in the real world, will define this research.

The Theory of Division of Moral Labour as a Tool to Understand the Lure of AI Use in Counter-Terrorism

As this research focuses on the efforts made by counter-terrorism agencies, a significant part of the discussion will be dedicated to counter-terrorism and its effectiveness. CTS scholars have invested significant effort in analysing the under covered intentions, effectiveness and impact of counter-terror methods at the international and national levels. The cost of the 'war on terror', in terms of human suffering and from a financial and political point of view, has been examined, for example (Jackson 2005; Khan and Kaunert 2023), as has been the impact on individuals

[4] The term 'Jew' is, of course, also a social construct.

belonging to a suspected group (Puar and Rai 2002). Counter-terrorism, according to CTS scholars, should be researched, just like terrorism, by understanding the whole political and sociological impact of the countermeasures in order to avoid what Zulaika has identified as the pendulum of counter-terrorism, which are cases where counter-terrorism efforts merely bring more violence, which in turn leads to more counter efforts (Zulaika 2009). The critical theories of terrorism and counter-terrorism help in understanding and analyzing questions related to political power, violence, and the effectiveness of measures used by government counter-terrorism agencies. One of the key questions that Critical Terrorism Studies (CTS) continues to raise is: what other interests are involved in the choice of specific counter-terrorism measures? In the context of this research, the question is why counter-terrorism agencies are putting increasing effort into developing AI-based, person-focused predictive tools. On the surface of things, the answer to this question is quite simple, the current hype around AI technologies is leading every advanced organisation to see how they can incorporate those technologies (Floridi 2024). Counter-terrorism agencies are no different in that sense, especially as some of them perceive themselves as leading the way in adopting new technologies (Cornish 2010). However, although substantive organisational and financial efforts have been invested in creating counter-terrorism person-based predictive algorithms, there is still no significant evidence on their effectiveness. Even more so, most of the existing research claims that there are too many challenges involved in using the tools at this moment for them to be used at all (Ferguson 2017; Babuta et al. 2020; Wall 2024). Having said that, the hunt for such tools is still considered the 'holy grail' for counter-terrorism agencies (UNCCT 2021, 24). This research will show that the use of AI by counter-terrorism agencies to predict individual attackers has the potential to provide

agents with more than just a tool to reduce their manual work, but also a tool that shares the moral burden in attempts to conclude who could be a potential attacker. Therefore, the following paragraphs will discuss the theory of division of moral burden in society in general, and the translation of this theory to the relationship between AI and human beings and, in particular, AI and humans in matters of security.

Division of Moral Labour in Liberal Societies

The concept of the 'division of moral labour' is a central theme in contemporary political philosophy and ethics. At its core, the theory proposes that moral responsibilities are distributed between individuals and institutions, allowing each to fulfil specific roles in the maintenance of a just and ethical society. Samuel Scheffler, based on his reading of John Rawls writings, was one of the primary contributors to this theory, arguing that, by assigning the task of ensuring justice to institutional actors, individuals are free to pursue personal relationships, projects, and non-institutional moral obligations (Scheffler and Munoz-Dardé 2005). This theory suggests that living in a modern society creates a vast range of moral challenges for the individual in terms of solidarity with other members of society. A moral person, for example, cannot ignore issues such as hunger, homelessness, mental health and violence suffered by other members of society. However, addressing all of those issues all of the time will prevent him from being a productive member of society. The solution to this situation, according to Rawls, is the creations of institutions, such as state and tertiary sector organisations, which are dedicated to dealing with those problems, and the existence of those institutions takes the moral burden of responsibility away from the individual (Rawls 1993). By paying taxes and donating to human rights NGOs, the

individual can feel they have done their share in helping those who need help and they can then invest their time in their own development, which will of course contribute to society.

This neo-liberal theory, despite its appeal to some, has been the subject of serious criticism, particularly regarding the implications it has for personal moral responsibility. G.A. Cohen, a prominent critic of Rawlsian justice, argued that such a division risks allowing individuals to abdicate their personal ethical duties. According to Cohen, justice is not solely the purview of institutions, it must also be reflected in the daily choices and behaviours of individuals. When people rely exclusively on institutions to uphold justice, they may tolerate or even contribute to unjust outcomes by accepting incentives that conflict with egalitarian values (G. A. Cohen 2009). Another critical voice, Iris Marion Young, emphasised the importance of a 'social connection model' of responsibility. She argued that structural injustices cannot be addressed solely through institutional reform but require active engagement from individuals who are enmeshed in social systems. For Young, moral responsibility is not extinguished by institutional roles but is distributed across all participants in a system that produces injustice (Young 2006). Lisa Herzog further critiqued the division by highlighting how organisational structures can obscure moral responsibility. In her work on ethics in economic systems, Herzog shows how individuals in bureaucratic or corporate settings may experience 'moral deskilling', a condition in which ethical reasoning is eroded due to routinisation and reliance on external procedures (Herzog 2018). Similarly, Albert Bandura's concept of 'moral disengagement' illustrates how diffusion of responsibility in complex systems enables individuals to distance themselves from the consequences of their actions and does so by presenting, among other examples, the case of US

counter-terrorism agencies during the ‘war on terror’ under the Bush administration (Bandura 2004).

These criticisms converge on a common concern: the danger that the division of moral labour can become a mechanism for moral abdication. Rather than empowering ethical cooperation, it may foster environments where responsibility is diluted, enabling actors, both individuals and institutions, to evade moral scrutiny. Regardless of how critical or supportive the research is regarding the theories of division of moral labour, there is a general agreement on this phenomenon, as an observation, that individuals are looking at societies and organisations as an outlet for their moral burden that increases the more they are aware of the injustices that exist in the societal structures they are part of.

Moral Delegation in the Age of Artificial Intelligence

The criticism of delegating moral responsibility becomes even more pressing with the rise of artificial intelligence (AI). As AI systems are increasingly deployed to make decisions in domains such as healthcare, criminal justice, warfare, and social services, a new form of moral outsourcing is emerging. This involves the transfer of decision-making authority - and by extension, moral responsibility - from human agents to algorithmic systems. The first question that arises from this ever-growing phenomenon, is whether we can perceive algorithms as separate ‘agents’ from their designers and whether we can attribute any ‘moral agency’ to the decisions of the algorithm. This issue has been discussed in the growing philosophical writing on the moral roles of algorithms in our daily life (Fritz et al. 2020) and I would like to focus on three different approaches to this question.

Luciano Floridi (2001) suggests a model that is based on the ability to abstract the actions of a machine at a higher level. An algorithm should be perceived as a 'moral agent' not based on whether it possesses an 'intention' in its actions, but on whether we can abstract such an 'intention' from its results (Floridi and Sanders 2001, 58). Let's take, for example, a parole recommendation system such as that which is increasingly common in the US legal system (Doaa Abu Elyounes 2020). Such a system, which obtains data, calculates a result and offers a result based on continuous uploaded data, is an 'agent' following Floridi's logic even though it is invisible to us physically. What makes such an 'agent' a 'moral agent' is when we perceive the actions of the 'agent' as having moral implications. In this example, the fact that those algorithms suggest more convenient parole conditions to white people (Flores, Bechtel, and Lowenkamp 2016) makes the algorithms 'moral agents' regardless of whether or not we show where the 'intention' for such a moral action lies. The issue with such a model which makes those algorithms 'moral agents' is clear, it makes the algorithm responsible for it moral action, but it certainly does not solve the problem of its accountability to its actions, meaning if the algorithm is a 'moral agent' who is responsible for its moral decisions?

A different approach to Floridi's understanding of algorithms as 'agents' and then as 'moral agents' was adopted by Deborah Johnson (Johnson and Verdicchio 2019). According to her, agency can be casual, where the effect of an action is accidental for example, it could be intentional, when someone wants to achieve a goal and then act accordingly to get it, and there could be 'triadic agency', where the agency is a combination of human and machine actions in order to achieve a goal (Johnson and Verdicchio 2019, 4). If we look at the parole algorithm as an example, the judicial system wants a way to statistically assess how dangerous it would be to

release a prisoner on bail, a designer builds this tool, and the tool makes the assessment. The responsibility and the 'moral agency' lies, according to Johnson, only on the side that we can attribute intention to and, in this case, on the designer that intended their tool to provide an accurate design for the algorithm. By removing all responsibility from the algorithm, Johnson redacts the term 'agency', which she attributed to the algorithm in her 'triadic agency' theory and compares the biased actions of the algorithm to a hammer that was designed badly and therefore bends the nails.

A third approach to the 'moral agency' question in the relationship between algorithms and humans is the 'hybrid agency' approach as developed by Verbeek (2006). According to him, the separation between actions and moral decisions is artificial and, in the case of human and machine interaction, no human or object by itself has 'moral agency'. Instead, 'moral agency' is created by a combination of things (Verbeek 2006, 364). Returning to the parole algorithm example, the designer and the algorithm that comes to a final decision about bail are creating together a 'hybrid moral agency' of human and machine that should not be separated. Verbeek does not neglect the question of accountability, in his eyes the designers of algorithms are not only creating a moral agency, they are creating a whole new hybrid reality. Therefore, they should be morally accountable for the results of this reality (Verbeek 2006, 368). This approach, although very tempting in its holistic view of the interaction between human and algorithm is, in practice, very similar to Johnson's in its outcome as it discusses the roles of the algorithm and its designer and attributes all of the responsibility and accountability for the algorithm's moral choices to its designer.

A major shortcoming in all those approaches to the possible 'moral agency' of an algorithm is that they ignore the role and the 'moral agency' of the user of the algorithm who is, most of the time, not its designer. In the example discussed earlier, the question is, what is the 'moral agency' of the judge that receives a biased parole assessment from an algorithm and makes their judgment based on it?

The question of 'moral agency' or blame on the users of algorithms that have a moral impact in reality has been the subject of a few recent studies. The research is focused on two main questions: how do the users feel about their responsibilities when using an algorithm? And how is their responsibility is perceived by others, when they are using the algorithms?

Regarding the first question, on whether users of algorithms delegate their moral agency and how they feel about it, an interesting trial was recently conducted asking those specific questions (Salatino et al. 2025). Salatino and her team, created a test that checks the response of drone operators to difficult moral scenarios around using force in a mixed civilian and military area. The test required three groups to decide on whether to attack or not. One group was without an AI assistant, another group had the help of an aggressive AI assistant that suggested attacks in the situations, and one group had an AI assistant that suggested refraining from attacks in the situations (Salatino et al. 2025, 3). The results were very clear. First, the human operators trusted the AI recommendation, regardless of whether it was aggressive or passive and mostly followed its recommendations to attack or not to attack even in clear cut cases of violations of international law, on the one hand, or absence of civilian danger, on the other. The second finding was that the speed of the decision whether to attack or not, rose significantly when using an aggressive AI or not, compared to when it was just a human decision. And the third, maybe the

most interesting, finding was when presented with the outcome of their decisions, the operators that chose to attack based on the aggressive AI recommendations, felt much less responsible for the outcome, compared to those that chose to attack in the same situation without the AI assistant (Salatino et al. 2025, 8–11). This study shows that, even when it comes to very complex moral decisions, humans feel very comfortable delegating these moral decisions to an AI algorithm, both in real time during the decision making and in retrospect (i.e., after the outcome of the decision is exposed). Another study shows that it is not only the users that feel comfortable with shifting the moral responsibility to AI but also society in general (Feier et al. 2021). The study examined whether there was a change in the perception of the responsibility of a delegator between delegating a duty to a human or to a machine. It looked at two different situations - how the public views the delegator's responsibility when the outcome of the delegation was positive or negative. It found that, when the outcome was positive, there was no real change in perception between delegating the action to a human or to a machine. However, when the result of the delegation was negative, delegators were judged with much more leniency where the mistakes was made by an algorithm (Feier et al. 2021, 14). Thus, if you delegate your duty to a human that makes a mistake, you are much more responsible for the mistake than if you delegated the same job to a machine. This finding, as the researchers claim, might cause individuals and/or institutions to 'over-utilize artificial agents even when they are not the ideal choice for a given task, because they wish to use machines as scapegoats' (Feier et al. 2021, 3).

Most of the people in a healthy society wish to think of themselves as moral people. However, it is not a simple task to be a moral person and, one of the ways to deal with it, is to divide the moral burden with others. The rise of algorithms that are in charge of actions with a high moral

impact has made the division of moral labour easier, as the machine that has become the final 'moral agent' has no issue with carrying that responsibly. This creates a very enticing power - to use algorithms in moral complex situations, such as ones that are common in counter-terrorism organisations, and therefore an important prism for examining any counter-terrorism algorithm that these organisations are using.

Assessing the use of AI in counter-terrorism using Lindahl's Critical Theory Model of Counter-Terrorism

After discussing the possible reasons behind counter-terrorism agencies' rush to develop predictive AI algorithms, another question that Critical Terrorism Studies (CTS) might ask is: should such tools be used at all in a sensitive environment like counter-terrorism? In his book, *A Critical Theory of Counterterrorism*, Lindahl (2019) presented an ethical model of counter-terrorism that would break that cycle of violence and present a holistic solution to repeated violence. In order to explain how creating such a model fits with CTS theories, Lindahl uses Weber's 'ideal type' theory (Weber 1917). The 'ideal type' is a tool that gives meaning to the researcher's constructs of reality and allows others to compare it and use it based on its definitions; it is not the utopian version of reality; it is the moral and ethical structure of the existing reality as perceived by the researcher. Lindahl's model, therefore, is a social construct of the existing realities of counter-terrorism implemented utilising the values of CTS. It is not based on a specific existing counter-terrorism model but is used solely to critically analyse different counter-terrorism strategies. Some of the components of Lindahl's counter-terrorism model might sound utopian when discussing existing counter-terrorism strategies, such as adopting the principles of non-violence and emancipation. However, it is important to acknowledge that the

model is a tool to check and compare counter-terrorism strategies, so the real question when working with the model is how close any particular counter-terrorism strategy is to the model. Applying the model will check, for example, how much the counter-terrorism strategy tries to incorporate non-violence measures and whether this strategy is entirely non-violent.

Lindahl's Counter-Terrorism Model

Lindahl's model of counter-terrorism consists of five elements, each evaluating a different stage in the creation of a counter-terrorism strategy. In his book, he examines the Norwegian counter-terrorism strategy to present a possible use of the model.

The five elements of the model are:

Key Assumptions - This component suggests that any counter-terrorism strategy must first acknowledge the challenges of defining the terrorism that needs to be countered. This process includes the understanding that terrorism is a social construct and that acts of terrorism are always rooted in some background that can usually be found in deep politics. That understanding is necessary to define both the problem and the possible answer (Lindahl 2019, 92). When Lindahl assessed the key assumptions of the Norwegian counter-terrorism strategy, he concluded that, regarding international terrorism, Norway has been operating under the assumption that it is one cohesive international problem without any attempt to dive into and map the different reasons and players that might affect Norway, and the connection to the deep politics of Norway (Lindahl 2019, 138–39).

Basic principles - This component in the model includes the five basic principles of the counter-terrorism 'ideal type' model. According to Lindahl, checking if those principles were applicable to a counter-terrorism strategy is the basis of applying the model. The principles are:

Dare to know: This principle checks if the counter-terrorism strategy includes signs of 'a commitment to explore and question the knowledge and assumptions we already hold about terrorism' (Lindahl 2019, 93). When examining this principle in the Norwegian counter-terrorism strategy, Lindahl finds genuine attempts to ask questions following the Breivik attacks, together with some avoidance of a discussion on structural causes. This makes him conclude that there is a 'lite dare to know' principle in the Norwegian counter-terrorism strategy (Lindahl 2019, 141).

Emancipation: This principle requires that a counter-terrorism strategy holds a normative position, acknowledging that gaining security cannot be done at the cost of other people's security. Lindahl understands that finding an example of a complete adaptation of the emancipation principle will be hard. However, a clear normative position can be found in some; in the case of the Norwegian counter-terrorism strategy, Lindahl finds the normative framework in the Norwegian commitment to international law and specifically international human rights law.

Means/Ends: This principle requires that any counter-terrorism strategy consider whether the means it uses are consistent with the result required from any counter-terrorism strategy, the end of the cycle of violence. The meaning is that any strategy that in the short term might reduce the violence but in the long term might create even more violence is not fit to end. When Lindahl reviewed whether this principle applies to the Norwegian counter-terrorism strategy, he acknowledged that Norway's strong commitment to human rights made most of its tactics fit the principle. However, he suggests that Norway's involvement as a NATO member in NATO's military operations contradicts it (Lindahl 2019, 145).

Non-Violence: In the Lindahl model, this principle is probably the most challenging one. According to Lindahl, any kind of violence, in any circumstances, is an ill fit for countering terrorism. Adopting the idea of non-violence, as both a practical and a moral choice, could, according to Lindahl, lead to a push for better preventive tactics that will not involve violence. For example, he presents the Norwegian approach of keeping dealing with terrorism in the hands of the local police under the laws that they are bound to, and not giving it to special agencies, with special permissions to use violent tactics (Lindahl 2019, 146).

Holism: This principle of holism in the model is meant to check if all the previous principles are considered together as a whole. The idea is that only by combining the principles of dare to know, emancipation, means/ends and non-violence can you achieve a holistic approach to counter-terrorism (Lindahl 2019, 97).

Strategies and Tactics - This component recommends 'ideal type' counter-terrorism strategies and tactics. The model, as Lindahl explains, cannot suggest and elaborate all the possible strategies and tactics available to use against terror. Therefore, it suggests a general framework. The component describes two stages of dealing with a terror attack: the first stage, which is the immediate response to an attack, and the second stage, which is the long-term strategy response. For the first stage, the immediate response, Lindahl suggests that the main aim should be restoring life to normal as fast and efficiently as possible. To do that, Lindahl suggests adopting the natural disaster model of counter-terrorism. According to this model, because many times the first impact of terrorism is similar to a natural disaster (i.e., dead and injured people, damage to property, chaos, etc.), special rescue and health teams should be trained to be activated in the case of a terror attack. As for the later response, the component suggests that, although there

are a wide range of possible responses, all depending on the specific attack, those should be considered according to the model's basic principles, therefore avoiding initial overreaction, exploring the source of the attack and devise a reaction that could lead to the end of the violence, not encouraging more (Lindahl 2019, 97).

Priorities – This component emphasises that the primary role of any counter-terrorism strategy is the prevention of violence. Prevention of violence, according to Lindahl, should not be measured in how many terrorist attacks we have successfully managed to foil in the last year; rather, it should be measured in the influence that proactive counter-terrorism had on lowering the deep issue tensions that are the source of terrorism. Adopting strategies and tactics using the lens of conflict resolution theories is an example of such proactive prevention counter-terrorism measures (Lindahl 2019, 99). Examining the priorities component of the Norwegian counter-terrorism case study, Lindahl acknowledges the Norwegian commitment to proactive prevention as he identifies the efforts to prevent the conditions that will bring new violence. However, he resurfaces the question regarding the use of military operations in international counter-terrorism operations, which he sees as a potential measure to increase violence (Lindahl 2019, 149).

Evaluation - This component proposes that the evaluation methods for any counter-terrorism model should include a combination of proportionality, effectiveness and legitimacy as defined by Jackson (2011, 244). The proportionality and effectiveness of counter-terrorism measures should be reviewed using empirical tools. However, there are many questions regarding the right way to calculate those. According to Lindahl, many counter-terrorism organisations use a maximalist approach, meaning attributing a successfully foiled terror attack to any intervention

by them, regardless of the stage and level of seriousness of the foiled plan, and attribute full potential damage to any foiled attack, creating an exaggerated evaluation of the number of lives saved and economic damage prevented. This is done to justify the amount of resources agencies put into their prevention tactics. According to Lindahl, the solution to this problem is to try and evaluate those based on what we know rather than the worst-case scenario (Lindahl 2019, 103). The legitimation aspect of the evolution of counter-terrorism is checking if the effectiveness and proportionality were achieved in a legitimate way. According to Lindahl, a good way to examine this is to see if the actions were performed according to the guidance of human rights law.

In the above paragraphs I have reviewed at length Lindahl's model of critical counter-terrorism studies, which suggests a way to critically assess counter-terrorism strategies. However, this research is focused on a specific counter-terrorism tactic, which is the use of predictive AI to identify an individual before it decides to attack. It is obvious that only some parts of the Lindahl model are relevant when examining a specific counter-terrorism tactic. In the next chapter, I will present the research methodology based on these theories and Lindahl's model and will explain how and when I use Lindahl's model during the research.

Conclusion

In this chapter, I discussed the theories relevant to this research, from my position as a researcher, that has led me to adopt critical perceptions of terrorism and counter-terrorism and Lindahl's model of counter-terrorism in particular, up to the understanding of the potential lure of using algorithms in counter-terrorism using the theoretical framework of division of moral labour and the moral impact of applying algorithms for moral actions in our day-to-day life. In the next chapter, I will discuss how those theories have impacted the design of the methodology.

Chapter 4 - Research Methodology: From Case Selection to Data Analysis via the Israeli Supreme Court

In this chapter, I start with a discussion on the choice of the case study and the impact of the theoretical framework on the research methodology. The second part of the chapter describes how the data was collected, focusing more on the challenges that arose following the freedom of information request that reached the Israeli Supreme Court and in obtaining the interviews with former ISA agents. The last part of the chapter discusses, more broadly, the methodology of the research, the analysis that was done to obtain as many observations as possible on the available data - with a particular focus on indictments against Palestinians that came about as a result of the AI tool - and a discussion about the applicability of using Sondre Lindahl's critical terrorism studies model of counter-terrorism as a way to assess a predictive AI tool that could identify individual attacks.

The ISA AI Tool as a Single Case Study

The empirical data used in this research was based solely on the data collected from the Israeli ISA AI case study. The main reason for that is that, during the design part of this research, I could not find any other similar relevant case study with usable empirical data. As explained later in this chapter, a unique set of circumstances has led to the fact that this research collected empirical data on the case study as, usually, counter-terrorism agencies do not share any data regarding their tools. Because this research was focused on the ability of predictive AI to predict a singular activity like the decision to commit an individual attack, I did not want to mix other case studies from person-based predictive policing AI that might have some empirical data but which were focused on much more common offences and therefore raise different questions.

Having said that, the choice to use a single case study in exploratory research like this is not exceptional when no other case studies are available, the data collected includes particular information (Stake 1995; Yin 2018), and when the case study can contribute to theory and practice (Flyvbjerg 2006), as I hope this research can. However, when working with a single case study, there is a need to be extra careful regarding the way you collect analysis and treat the data, especially when, as in the case of this specific case study, you were personally involved in it.

Positionality, Critical Terrorism Studies, and the Methodology

Choosing a critical research approach requires a healthy amount of scepticism and reflexivity, as discussed in the previous chapter. The main task for any critical researcher is to challenge their preconceptions and to constantly ask if the methodology that has been chosen is impacted by those misconceptions and, therefore, affects the integrity of the research. Having been exposed to too many wrongdoings by the Israeli security forces, prosecution, and courts, my pre-position regarding those is suspicion at the very least. My challenge then, as a researcher during this research, was to question myself: do those pre-positions impact me in a way that leads me to ignore valid arguments and data the Israeli security forces brought up when explaining what they perceive as the success of their AI counter-terrorism tool? It is, in a way, an opposite challenge to the one CTS is raising as, for example, instead of immediately questioning and identifying the state and social mechanisms around 'what is terrorism?' and 'who is a terrorist?', my challenge was to question my beliefs and to try and identify if there was something in the constructs created by the Israeli forces that can be substantiated in real events. The same can be said regarding the abilities of artificial intelligence. Although I was very sceptical about the abilities of

AI to predict such unique human behaviour, I have tried throughout the research to challenge this with the relevant questions especially in light of the increased effectiveness of AI technologies in recent years. Like when a mathematician, to prove that the answer they have found is the only possible answer, must show that all other solutions are invalid. That is why this research starts by exploring if the Israeli AI tool was effective, while accepting the social constructs that defined it. Only after answering that do I examine it using deconstructivism tools. This challenge will define the methodology of the research.

AI tools in counter-terrorism are relatively new, so this research is exploratory. During the first stage of the research, I tried to understand what led the ISA to decide to develop such a tool, its purpose, how it operated, and what the results were. All these questions were asked while simultaneously identifying that they were based only on the Israeli security services' definitions and measurements. This means that questions like, 'what is an attack?', 'who is an attacker?', 'what are the characteristics of an attacker?', 'what is radicalisation?', and 'what is success when it comes to counter-terrorism?' are all examined firstly by using the definitions the Israeli security forces used. There are a few reasons for this choice. The first is technical because the ISA created the tool and defined its goals and purposes. In order to understand what it was doing and why, it is vital first to examine it using the prism it was created through. The second reason is that all the collected data came directly or indirectly from the Israeli security forces; therefore, it should be analysed at the start in that way. The third reason is found in my initial pre-position that is sceptical of the possibility of such a tool being effective, even when using the definitions of the Israeli security services.

Another way to assess the AI tool's abilities and impact was to analyse the relevant Israeli military court indictment that followed the AI identification. The first part was to identify the specific AI-related indictment; the second part was to try and analyse what we can learn from the indictment, especially from the Facebook pages and posts connected to them, about what the AI tool was looking for. The second part was to follow the court's reaction to those indictments and their impact on the people indicted.

Using the data collected and its implications, a first analysis of the data was performed using all of the parameters established by the ISA to measure success. This analysis uses a Critical Terrorism Studies lens to first examine all of the initial definitions used by the ISA to create the tool. The second phase of the analysis is dedicated to exposing other motives that led to the ISA being so keen on developing and activating such a system, using the theory on division of moral labour in the context of human algorithm relationship. The third phase of the analysis is carried out in order to answer the question 'was the AI tool successful, according to the ISA's parameters?' After analysing the tool using the ISA's parameters, a further analysis was performed using more critical eyes, regarding the potential effectiveness of the tool in the short and long term. Using the findings and the analysis carried out on the ISA tool, the research suggests the conditions needed to create a theoretically potentially more effective AI tool. The research then focuses on analysing this theoretical tool using Lindahl's model of counter-terrorism, as detailed in the previous chapter. By applying the model to the theoretically effective AI tool that can potentially identify an attacker before its attack, the research tries to answer the question, 'should we develop such a counter-terrorism tactic at all, even if it could be effective?' In the last part, the research presents a different theoretical way to use AI algorithms to help

cope with the individual attack phenomenon - this time using it to predict waves of attacks not the individual ones - and it also analyses this theoretical tool using Lindahl's model of counter-terrorism.

Collecting the Data and Analysing it

The Conditions that Facilitated Obtaining the Data

This research explores the activation of a counter-terrorism AI tool by a secretive counter terrorism agency, so it is not performed in a data sharing environment. As explained in the Introduction, my interest in studying the ISA's AI predictive tool came from a personal acquaintance and interest when I was exposed to its impacts. Having said that, this research became more possible in terms of access to data because of the unique situation that applied in Israel and the OPT at the time. The application of a preventive AI tool requires surveillance of the population to such an extent that it can be done either in a non-democratic setting, for example, in China (Qiang 2019), or illegally in democratic settings, such as in the NSA case, as revealed by Edward Snowden (Reuters 2020). Either way, the ability to access data on surveillance programmes is very limited, as security agencies are generally reluctant to expose data, but even more so when it comes to programmes operating in a non-democratic setting or acting illegally. A relatively rare combination of conditions allowed me to access some data about the ISA's AI tool. The first condition that allowed me access was Israel's unique legal situation regarding its citizens and the Palestinians in the OPT. The continuous discussion regarding whether Israel is running an apartheid regime (Amnesty International 2022) is based mainly on the fact that, whilst under military law in the OPT, Palestinians are deprived of the most basic human rights, including privacy; Jewish Israeli citizens enjoy relatively liberal democratic rights.

It is this unique legal situation in Israel and the OPT that, on the one hand, allowed the ISA to act legally under the military law regime when conducting a massive surveillance programme on Palestinians and ignoring any question regarding their rights and, on the other, under the Israeli legal system, allowed me to file a freedom of information request and bring the ISA to court in order to comply with my request.

Another condition that helped obtain information on the ISA's AI tool was that, for reasons that will be discussed later in the research,[5] the ISA eventually allowed several people involved in the creation and application of the ISA tool to publish articles and to be interviewed about the tool and its success (Berbing 2019; Shahaf 2020; Y. S. 2021). Those articles included very important information regarding the AI tool, but they were also helpful in my later attempts to interview ISA agents about the tool, as its existence was already a known public fact, which meant that the interviewees were less reluctant to discuss it.

Generally, the research data was obtained using a freedom of information request, interviews, and open-source materials, and each required a different approach to obtaining and processing the information.

The Scope of the Gathered Data

Before starting the research, the initial information I had was based on a combination of media articles, the ISA publication on the tool as mentioned above, and my personal experience from the military courts. From these sources, I knew that the Israeli security forces referred to the specific wave of violence that led to the implementation of the preventive AI tool as 'Operation Godel Hashaa', and this is dated to the period between October 2015 and September 2016

[5] See chapter 7.

(Carmeli 2019). I also knew that the AI tool was activated around January 2016, as this was the first time we, as lawyers, had heard about cases based on it. This basic information led me to focus my data gathering on the period between March 2015 and March 2017. This time frame was chosen as it includes four time periods: first, the period before the researched wave of violence started; second, the period after the violence started and before the AI tool was created and activated; third, the period after the AI tool was created and activated during the wave of violence; and fourth, the period after the wave of violence had ended but the AI tool was still active. Comparing those four periods allowed me to try and identify the impact of the ISA's AI tool, if such a tool existed.

In terms of the data I was looking for that was not open source, I firstly wanted to have all the statistical data regarding what the Israeli security forces defined as terror attacks, by months, by kind of attack, and by results. This data I knew existed in the ISA statistical monthly reports. Those reports were sometimes quoted as data by the ISA in Israeli media (Hasson 2014) and were not classified. However, they were not accessible to the public, and I knew I needed to file a freedom of information request to access them. Other information I wanted to obtain access to was regarding statistical data and specific relevant indictment issues in the Israeli military courts' indictments against Palestinians. The Israeli military courts were the legal tool through which the ISA arrested and charged those who were identified as extremely dangerous, according to the AI tool recommendations. The statistical data from the military courts regarding the number of indictments was needed to identify the impact the AI-based indictments had on the courts' activity. The indictments themselves were needed for two reasons: first, to identify the specific cases based on the AI tool, and second, to try to identify the signals the AI tool was looking for in

the social media of the people identified as potential attackers. As will be explained at length in Chapter Six, the indictments in the Israeli military courts include the charges against each defendant, which includes the legal article and the factual description. In the cases against the people who were identified by the AI tool, the only factual description was the defendant's social media posts screenshots which were attached to the indictments. This data was also defined as public information and was also inaccessible to the public and, therefore, was also a part of the freedom of information request. Apart from this data, I wanted to interview ex-agents of the ISA who worked on the project to fill in the gaps in the available information regarding the tool as much as possible. I wanted furthermore to interview some of the Israeli military court judges who took part in the cases that were initiated by the AI tool, in order to understand their perspective. However, my interactions with the military court system over the freedom of information request, as will be detailed later, prevented it. At the beginning of the research, I was wondering about interviewing Palestinians who were arrested because the AI system identified them, but as the research scope and question became clearer, I decided to forsake this idea as the ethical risks involved in conducting such interviews were higher than the benefit to the research.

The data collection period for this research was from 2020 to 2023. During this period, the whole world was also dealing with the COVID-19 pandemic. The different approaches to dealing with the pandemic between Ireland and Israel made travelling to Israel a challenging quest. I could not visit Israel during the whole of 2021. However, once vaccinated, I could continue my research easily as Israel was one of the fastest countries to open fully.

To summarise, the data used for this research was collected through a freedom of information request, interviews with ex-ISA agents, and open-source materials.

The Freedom of Information Request and Legal Proceedings

The attempt to obtain information from the Israeli security forces using a freedom of information request was ultimately successful but required considerable time and effort. The following paragraphs describes the process in chronological order.

The Israeli 'Law of Freedom of Information' was enacted in 1998.[6] It includes several mechanisms regarding freedom of information; the relevant ones to this research are Article 5, which states that every public authority is required to publish a yearly report about its activities and responsibilities, and Article 7, which defines the terms under which a freedom of information request can be submitted to a public authority and should be answered. The law exempts the ISA from answering requests according to Article 7, but there is no exemption from the duty to publish statistical reports, according to Article 5.

The specific data I was looking to receive was public data that has not been published and included:

a) The statistical reports the ISA was supposed to publish regarding the relevant period, according to Article 5. These include the number of attacks, the kind of attacks, the results of attacks, and a summary of each month's trends. As I already mentioned, I knew those reports existed as the ISA referred to them in its press releases.
b) From the Israeli military courts, I wanted to receive the number of cases in the Israeli military courts per month and their distribution per month according to the offences in the relevant period.

[6] Book of Laws 1667 5758 p 226.

c) As for the indictments themselves, the requested documents were all of the indictments that had been filed with the Israeli military courts during the requested period that either included charges that might be used for the AI tool cases, such as: planning and conspiring to commit attacks, attempted attacks and incitement to violence, or indictments that were filed during that period which carried the relevant keywords 'social media' 'Facebook, or 'Twitter' since, according to other information that had been gathered up to that point, the indictments that were filed against people identified by the AI tool had focused on their online activity.

Before filing the freedom of information request, I first tried to receive some of the information directly from the Israeli military courts and the ISA in an informal way. Israel is a country that appreciates personal connections, and informal requests to people you know personally might be more effective than formal requests. On 5 January 2021, I called one of the ISA legal advisors to the military courts, whom I knew, and asked him if they could publish the statistical reports. I was told there was a plan to launch a new website for the ISA that would include those reports, but it was made very clear that the timeline for such a website was unclear. On 7 January 2021, I called one of the military courts' administrative officers, whom I also knew personally, and asked him about publishing statistical data regarding the cases (I knew that asking for the indictments themselves would be pointless, following previous interactions regarding those). I was told by him that the president of the military courts had decided not to publish those following criticism the courts received when they published those statistics in 2015 from 'you and your leftist friends', to use his words. Following these interactions, I filed an official freedom of information request with my previous law office, Gaby Lasky and Partners, with the Israeli army authorities

via the office of the Israeli army spokesperson on 4 February 2021. The request asked for all of the information detailed a.) to c.) above.

The request was filed according to Israeli freedom of information law, which also sets out that the authority needs to provide an answer to the request within 30 days on whether they are going to comply with or decline the request according to the different reasons for declining that are mentioned in Article 8 of the law. As expected, the army representative did not reply on time. On 6 April 2021, I received their initial response, which included all of the requested statistical information. This included the ISA's monthly and yearly activity reports comprehensively documenting what the ISA identified as hostile terrorist activities per category per month and a short trend analysis paragraph.[7] The statistical data also included a breakdown from the Israeli military courts relating to the number of indictments filed monthly according to the main charge mentioned.

As for the indictments from the Israeli military courts, the army denied the request, claiming that 'locating the requested information requires an individual examination of thousands of files, in a way that amounts to an unreasonable allocation of resources. Therefore, we will not be able to provide the requested information'. The IDF spokesperson chose to use the phrase 'unreasonable allocation of resources' as it is one of the conditions that allows a public authority to deny a request according to Article 8(1) of the Law. On 14 April 2021, I replied to the IDF spokesperson, claiming that the jurisprudence regarding the interpretation of 'unreasonable allocation of resources' in Article 8(1) is that a public authority cannot make such a general claim

[7] On May 2023, the ISA launched a new web site that includes all monthly reports since March 2014. Those can be found here: https://www.shabak.gov.il/reports/

and it needs to specify the specific efforts that are required and what makes them unreasonable. I also offered to receive the thousands of relevant files and review them myself. While waiting for an answer from the military spokesman, on 10 May 2021, another round of violence erupted between Palestinians and Israelis, named 'Operation The Sword of al Aqsa' by Palestinians and 'The Guardians of the Walls' by Israelis. This resulted in hundreds of deaths and the severe destruction of property. As with any round of violence and its aftermath, the IDF spokesperson's attention was on the ongoing events, and there was no chance of receiving any cooperation from them regarding the freedom of information request. Only on 21 July 2021, was a short reply received from the IDF spokesperson claiming that:

> 'A reply to your request will require the searching and collecting of 6,000 indictments, and an individual examination of each one of the indictments, including the content of the indictments, the identity of the defendant, and the limitations of passing it on, according to the law, or according to a court decision. The process of collecting and examining those indictments represents an unreasonable allocation of resources. Therefore, according to Article 8(10) of the law, we cannot respond to your request.'[8]

The only possible solution left was to file an administrative petition to the Israeli District Court according to Article 17 of Israeli freedom of information law. Thus, on 24 October 2021, a petition was filed (case no. 55892-10-21 Ramati Vs Israel Defence forces – Jerusalem District Court).[9] The petition focused on the military claim that the indictments cannot be delivered due to the

[8] A letter from the IDF spokesperson on 21/07/2021, in the author's collection.

[9] All of the legal proceeding were carried out with the help of Adv. Karin Torn of Kalai Rosen and Co., Partners in Law, Law Offices.

unreasonable allocation of resources needed and stated that, because the Army has a computerised system where all of the indictments are collected, it is very easy to pull the relevant indictments. Following the filing of the petition to the court and even before any hearing had been scheduled, on 3 January 2022, the army transferred 90 indictments with 'incitement to violence' marked as a primary offence for the relevant period. In the response, the army claimed that no indictments carried the offences of planning and conspiring to commit attacks or attempted attacks as a primary offence during that period. This was the first time the Israeli army had agreed to hand over in bulk, for research purposes, indictments filed at the Israeli military courts. However, the army still refused to supply the indictments according to the requested keywords (i.e., Facebook, Twitter, social media). This was necessary to identify if any of the indictments used against people identified by the AI tool chose to use a different main charge than the one asked in the request (planning and conspiring to commit attacks, attempted attacks and incitement to violence). During the hearings in the District Court in Jerusalem, the army continued to argue that their computer system was outdated and did not allow them to look for specific keywords. Therefore, the only way to locate the documents was to review all 6,000 indictments filed during that period manually. Luckily enough, the District Court judge was sceptical about the army's answer and agreed to hear evidence, a generally rare occurrence in Israeli administrative proceedings, where the courts usually accept factual statements by the state and the hearing is based more on the legal arguments. The testimony of the military courts' technical officer on 16 June 2022 was key as it became clear that there was a technical way to filter the military court database according to keywords.

On 22 November 2022, the District Court ruled and ordered the army to deliver the requested documents to me in 60 days. The court ruled that although retrieving the data requires some irregular efforts from the army, the importance of receiving the information is greater:

> 'The respondent may be required to carry out information retrieval activities in such an examination, even if it is exceptional, within the framework of the "filters" of "reasonableness" and "proportionality" in Sections 10 and 11 of the Law, which the Authority must apply in the context of balancing interests, when considering a refusal to provide information under Sections 8 and 9 of the law. This is requested in light of the petitioner's interest in the information as part of his academic research and doctoral thesis and in light of the great public interest in academic research, due to its contribution to society in all areas of life, including the field of law, and on the issue of the influence of social networks in general, and on terrorism and radicalisation in particular. For this reason, given the importance of the information to the applicant, and even to the public, the exceptional information retrieval activity is required.'[10]

On 22 January 2023, the last day for receiving the documents, the army filed an appeal against the District Court ruling to the Israeli Supreme Court (case no. 637/23 Israeli Defence Forces Vs Ramati, Supreme Court of Israel). The appeal against the District Court hearing was filed this time by the Israeli State Attorney's office and included completely new arguments. This time, the main argument focused on the fact that the army was not required to submit indictments from the Israeli military courts following a freedom of information request, probably based on the justified

[10] Case no. 55892-10-21 Ramati Vs Israeli Defence Army – Jerusalem District Court. Available at https://www.nevo.co.il/psika_html/minhali/MM-21-10-55892-435.htm

fear that the District Court decision would open the door to others requesting information from the Israeli military courts. The most outrageous argument raised by the state in their appeal was that handing over the indictments would hurt the right to privacy of Palestinians, the same right that was so carelessly ignored when obtaining the information that was the basis for the indictments.

Following a hearing in the Israeli Supreme Court on 3 August 2023, the representative of the Israeli State Attorney recognised that the court would reject their claims, mainly because most of their arguments were new arguments not mentioned at the original hearing in the District Court. In order to avoid a Supreme Court decision agreeing with the District Court decision, the Israeli State Attorney's office decided to withdraw the appeal and on 7 December 2023, almost three years after my initial request, I received 254 indictments, which included the search words Facebook, Twitter, and social media as requested in the initial freedom of information request, the court also ruled that the state will pay the law firm that represented me the expenses of the legal process. The Israeli State Attorney's choice to withdraw from the appeal left the District Court decision as a standing precedent and, as of that moment, the only jurisprudence regarding the duty of the Israeli military court to produce indictments for research purposes. The court choice to justify its ruling by emphasising the importance of academic research, created a precedent that will allow other researchers to receive bulk data segmented by categories upon request.

The Interviews

Apart from the interviews, all the information gathered in this research was public and, therefore, did not raise particular ethical issues. The only issue that required ethical approval and attention were the interviews with the ISA agents. I applied for ethical approval of the research from the DCU Research Ethics Committee on 20 January 2022, focusing on the safeguarding of the interviewees. The application included a commitment to saving the interviews in a biometric password-protected computer and presenting the interviewees with an approved plain language statement and a detailed informed consent form. The application also stated that the planned interviewees were highly experienced Israeli intelligence officers, putting them at a very low likelihood to say things that might create a risk to them or others in the interviews. On 14 March 2022, I received approval for my research from the DCU Research Ethics Committee and could start with my interviews.

As previously explained, the ISA, being a counter-terrorism agency, is very reluctant to expose information and although it uses mass media, as all agencies do, the chances of them agreeing to or needing to cooperate with external academic research is low and, when it comes to someone who is openly criticising them, it is even lower. At that time, the ISA was already not happy with me because of a series of interviews I had given to the media following an article criticising their interrogation techniques (Ramati and Torn Hibler 2021) and about their questioning techniques and because I had forced them to publish their monthly reports (which they were obliged by law to publish). Therefore, I knew that the only way I could obtain direct information on the ISA tool was to interview retired agents who had been involved in creating the AI tool.

From my years working in the Israeli military courts, I knew a few ISA agents, especially those called by me to testify as witnesses in an attempt to expose wrongdoing in ISA investigations. Although those hearings would usually have become quite heated, with time, a particular kind of relationship began to form between me and some of the repeat witnesses, mainly during coffee breaks and long waits for court sessions where the schedule was always late. Active ISA agents are never presented by their real name but only by an alias that stays with them as long as they serve in the ISA, so I never knew their real name. However, I had the mobile number of some of them saved under their alias name. Just as Israel was starting to come out of Covid-19 lockdown in March 2021, I used those connections and reached out to several retired ISA agents I knew to interview them. Most of them did not respond, and three of them returned with some questions. One of them, BC, was a regional ISA commander of an area around Ramallah. He had been selected to help evaluate the outputs of the AI tool when it began. He introduced me to DB, a higher-ranking ISA analyst involved in creating and activating the AI tool from the start. After some negotiations, including their choosing to remain anonymous, they agreed to meet up and be interviewed on 26 April 2022 and 28 April 2022. I met BC at his home, and because I knew him, the conversation was very much open and friendly. It was just as Israel was coming out of an extended COVID-19 lockdown, and he, being retired, felt the need to speak. The interview took two and a half hours and included a lot of irrelevant gossip about people we both knew. BC was very helpful, but as he was lower ranked and had only started working with the AI tool after it had been created, his input was more limited to that stage and the general situation in the ISA during that period. The interview with DB was more structured and less generous. This was expected since he was a senior member, had worked with the AI team from the start, and did

not know me. He hosted me in his office at the new cyber company he was working in. He allocated us an hour, although, in the end, the interview took an hour and a half. He was much more thoughtful before answering the questions.

The fact that both of DB's commanders during the AI development had been giving interviews to the Israeli media regarding the AI tool (Shahaf 2020; Harel 2019) helped, as I could frame the questions based on their interviews, which made him more comfortable answering. Having said that, he was very careful to only elaborate on issues that had already been exposed, and to emphasise when he was saying things that were his own opinion. I felt the anonymity allowed both of them to speak more freely as, a couple of times before they said something, they asked to make sure that the interview was anonymised. Although the atmosphere in the interviews was very different, it was clear that both were very experienced intelligence agents as, for example, neither of them mentioned the real name of any other ISA agents they had worked with, and they easily pushed back on any question that was leading to what they felt was secret information.

The interviews were semi-structured as I needed both interviewees to fill some specific gaps in the available data. However, generally, the structure was mainly chronological, as I asked about their involvement in each stage of the tool's development, testing, and activation, only interrupting when I had a specific gap about a specific issue or where there were inconsistencies among the sources.

My original plan was to try to interview someone from the military intelligence technological unit (8200) of AMAN, the intelligence corps of the Israeli army, as they were involved in creating the tool and collecting some of the data fed into it. That proved challenging as I had never had any

interaction with this unit during my work and as they tend to be even more secretive than the ISA. The persons who serve in this unit are usually quickly incorporated into the Israeli high-tech industry, so I had to reach out to people I knew in that industry and ask them to refer me to ex-soldiers who had served in that unit. Most of the people I contacted refused. However, I managed to schedule another interview - with an ex-soldier who had been involved in creating the tool - for 15 October 2023. Eight days before the interview, the war between Hamas and Israel began. Like many reserve Israeli soldiers, he enlisted and stopped replying to my messages.

I also planned to interview military court judges regarding some of their rulings in AI cases. However, the freedom of information case regarding the indictments from the military courts had taken its toll. According to the state attorney prosecutor who was leading the case in the Israeli Supreme Court, the driving force behind trying to stop me from receiving the indictments was one specific personal in the Israeli military courts at the time, who also had the authority to approve military court judges to interview. The bitter ending, from that person point of view, of the case on the freedom of information request had a clear impact on my request to interview military judges. When I approached him directly when we met during the Supreme Court hearing on 3 August 2023, his amusing and direct answer was: 'in your dreams'.

The Open-Source Data

The open-source data included a few different kinds of data sources. The most direct and important open-source data were, as already mentioned, the articles by, and interviews with, ISA and military personnel who had dealt directly with the AI tool during the period relevant to the research. First to be recognised are two different Israeli security forces journals who each dedicated a separate issue to the challenges of the period I was focusing on and the solutions

provided by the different security agencies. Issue 4 of 'Methodology Insights' from April 2018 dedicated its 300 pages to Israeli intelligence efforts to collect and control digital information, and issue 22-23 of 'Routine Security' from October 2019 dedicated 220 pages to the period of the research and challenges it had presented for the Army and the ISA. On top of these very detailed journals, which also included articles directly relating to the ISA's AI tool (Y. 2018; Berbing 2019), there were also a series of interviews with the leaders of the ISA cyber unit, explicitly discussing the creation and activation of the AI tool (Harel 2019; Shahaf 2020). In 2021, a book by the head of the AMAN technological unit 8200 dedicated a whole chapter to the creation and impact of the tool (Y. S. 2021).

Another important source of open-source data were the Intelligence and Terrorism Information Center's (ITIC) weekly and special reports.[11] Those reports included statistics and the most recent intelligence analysis, as perceived at that time by the leading intelligence bodies. The third open-source data source used was the yearly statistical reports published by the Israeli military courts (Judea and Samaria 2016) which included statistical data about the number of cases per year per main charge. Table 1 summarises the open-source data used for the research.

[11] The ITIC is a part of the Israel Intelligence Heritage & Commemoration Center (IICC) whose members are veterans from all branches of Israel's intelligence community – the Institute for Intelligence and Special Operations (Mossad), the Israel Security Agency (ISA), and Israel Army Intelligence (AMAN) and, according to DB and BC, highly aligned with the ISA. Accessible here: https://www.terrorism-info.org.il/en/

Table 1. Open-source data

Information type	**Source**	**Description**
Information directly dealing with the application of the ISA AI	Barbing, Arik, and Or Glick. 2019. '(Heb.) Lone Terrorism – the ISA in "Operation Godel Hashaa."' *Routine Security*, The Campaign Between the Wars, no. 23, 127–48.	An article written by the former head of the AI programme in the ISA which describes the need, development, and activation of the AI tool.
	Y., Colonel. 2018. '(Heb.) The Journey Towards Clarifying the Perception and Implementation of Intelligence and Operational Superiority in the Digital Era.' *Methodology Insight, Big Data and Intelligence* 2.	An article written by the former head of the technological unit of the Israeli Army intelligence on their contribution to the AI tool.
	Y. S. 2021. *The Human Machine Team: How to Create Synergy between Human & Artificial Intelligence That Will Revolutionize Our World*.	A book written by the same author that includes a chapter on the ISA AI tool.
	Harel, Amos. 2019. 'How Israel Stopped a Third Palestinian Intifada.' *Haaretz*, October 4, 2019. https://www.haaretz.com/israel-news/2019-10-04/ty-article/.premium/how-israel-stopped-a-third-palestinian-	Two lengthy newspaper interviews with the heads of the ISA AI programmes. Both describe their point of view on the success of the AI tool,

	intifada/0000017f-e355-df7c-a5ff-e37f99d30000?v=1655910039370. Shahaf, Tal. 2020. 'ISA cyber chief: "From looking at 70 likes on Facebook, I know more about you than you know about yourself."' YNET. November 27, 2020. https://www.ynet.co.il/articles/0,7340,L-5851279,00.html.	revealing information about the tool's sources and abilities.
General information regarding the army and intelligence response to the 'Godel Hasha' campaign	*Routine Security*, The Campaign Between the Wars, no. 23 *Methodology Insight, Big Data and Intelligence no* 2.	Those two issues show the perspective of several army and intelligence officers regarding the violent period which is the subject of this research and the means that were taken to stop it.
Reports regarding the intelligence evaluation of the situation during the period researched.	https://www.terrorism-info.org.il/en/	107 weekly and special reports regarding the current events from the perspective of the Israeli intelligence community.

Using the three primary sources of data - the freedom of information request data, the interviews, and the open-source data - I was able to get a picture, although, of course, not a complete one, of the ISA's AI tool and its impact.

Data Period and Data Analysis

In order to identify the exact impact of the ISA's AI tool, I compared the collected data over 4 periods between 2015 and 2017. The reasoning behind this choice is that the ISA identified October 2015 to October 2016 as the period during which the violence took place. Allowing six months before and after that period allows changes to be identified:

- First period: March 2015 – September 2015: The period prior to the outbreak of violence, according to the ISA.
- Second period: October 2015 – January 2016: The period during the outbreak when the AI tool had not been implemented.
- Third period: February 2016 – October 2016: The period during the outbreak when the AI tool was in use.
- Fourth period: October 2016 – April 2017: The period after the end of the outbreak of violence, according to the ISA.

The logic behind choosing these periods is based on the possibility of identifying very specific changes and impacts by comparing them. I chose to start the research period in March 2015 and end it in April 2017 so that the four periods would be similar in length.

The data analysis of this research follows the general logic of its methodology. As the first stage of this research is to try to examine, through the eyes of the ISA, the need, impact, and effectiveness of the AI tool, the first analysis will focus on the data gathered according to the definitions of the Israeli security forces and the ISA especially. The analysis focuses on understanding the following trends:

1) Fluctuations in the patterns of the attacks and the attackers, according to the different periods. This will be done using statistical data about the number of attacks and attempted attacks, number of attackers per attack, method of attack, results of attacks, and the age, gender and geographical origin of attackers.
2) Data on the creation and activity of the AI Tool and its impact. This will be done mainly by using a combination of open-source publications, interviews, and a textual analysis of the indictments presented at the Israeli military courts that were based on AI recommendations.

Identifying and Analysing the AI-based Indictments

Following the freedom of information request, I received all of the indictments relating to the relevant security offences and keywords for the period. As the indictments themselves did not specifically mention being the product of an AI recommendation, the first stage of the analysis was to identify those exactly.

In Chapters five and six, I discuss at length the mechanisms that led to Palestinians being criminally indicted, based on the AI tool identifications. In a nutshell, the data collected from the open-source materials and the interviews pointed to the fact that when the ISA received a high-risk alert regarding a potential attacker, the agents chose one of three ways to react. One, the most common one, was making a call to the suspect or his family and warning them they were being watched, the second was to alert the Palestinian Authority security forces, and the third one, kept for the most dangerous cases, was to arrest and charge them in the Israeli military courts. The indictments against them used the incitement charge based on their social media activity prior to their arrest as they had not yet committed any other illegal act. The incitement

charge is very rarely used in an Israeli military court, so those indictments could have been easily identified; however, during the same period the AI tool was active, a team of soldiers from the Israeli military intelligence browsed the internet looking for popular Palestinian voices who encouraged violence against Israelis, and they were also arrested and charged with incitement. The challenge, therefore, was identifying which of those specific Indictments was created because of AI identification and which were because of work by people. Following an array of specific signals unique to the AI indictments (as elaborated in Chapter Six), I managed to identify 104 indictments against people identified by the ISA AI tool. I scanned all the indictments using a Hebrew OCR app and an Arabic OCR app, and all of the original posts attached to those indictments. After arranging and cleaning the data, I created an excel doc that included each textual post on a separate line. I identified 614 posts, 396 (64%) text posts and 218 (36%) visual posts (pictures or videos). I used this data to try to answer three different questions.

The first question was, 'what kind of signals was the AI tool looking for in the social media pages of the accused?' I manually went over the visual posts and categorised them according to their main signal: pictures of leaders, pictures of dead attackers, images of weapons, and Palestinian organisation insignias. To analyse the text posts, I used two Python scripts to identify commonalities. One Python script was a frequency script which scanned the text for the most repeated words in all of the posts and the most repeated combinations of words in each post up to a combination of 4 words. The second textual script I used is a semantic similarity script that groups posts with semantic similarity into clusters; by embedding an Arabic open-source semantic search (AraSAS) in the script, I clustered the different posts according to their semantic meaning. These two analyses helped me identify the most repeated keywords and topics in the

posts, which I used to analyse the specific signals searched by the ISA AI tool in the social media posts.

The second type of question I tried to answer using the indictments was, 'did any of the people indicted based on AI tool identification gave any indication that they were about to carry out an attack?' As the indictments in the military courts are supposed to present all the evidence there is against the defendant, such an indication in the form of a confession or other signals (suicide letter, a conversation with another) would have appeared in the indictments. To answer that question, I manually reviewed all the indictments and searched for such signals.

The third type of question I tried to answer using the indictments was 'what was the legal response to those indictments?' One of the challenges identified in the literature regarding the theoretical application of predictive AI is the legality of it (Wall 2024). I collected all the available jurisprudence based on the indictments and traced the legal analysis carried out by the Israeli military courts.

Answering those three questions, using the analyses of the AI-based indictments, together with all the other data collected from the freedom of information request, the open-source data, and the interviews, presents a relatively full picture regarding the creation, activation and operation of the ISA AI tool.

Analysing the Story behind the ISA's AI Tool and of Preventive AI Tools to Counter Individual Violence in General

In the second stage of the research, I examine the ISA's AI tool and the concept of a preventive AI tool in counter-terrorism using CTS theories. Firstly, I analyse the story behind the ISA's AI tool

and highlighted some of the ISA's blind spots regarding the initial definitions, data training, and creation of the tool. Then I discuss the ability to assess the effectiveness of such a tool, both in the broader sense - did the activation of the tool impact the wave of individual violence? - and in a narrow sense - does this tool pre-identify attackers before an attack?

In order to do this, I first used all of the statistical data collected and tried to identify if there was any statistical correlation between the number of individual attacks and the activation of the AI tool. At the second stage, I tried to find signals, using the indictments against those whom the AI identified, that can show intent to attack. This part of the analysis identified some weaknesses in the ISA AI tool, both in its creation and definitions and in the analysis of its effectiveness. I wanted to examine if there was a way to imagine a better tool that could avoid some of the identified difficulties in the ISA tool. I suggest how a theoretical tool could be more effective in pre-identifying individual attackers when operating in a democratic setting. For the final stage of the analysis, I use Lindahl's critical counter-terrorism model as a way of evaluating the suggested theoretical AI tool. I examine the tool through each of Lindahl's model elements, key assumptions, basic principles, strategies and tactics, priorities and evaluation and discuss whether we should aspire to develop such a theoretical AI tool at all.

The described two-stage analysis examines the counter-terrorism AI tool on two levels. The first level uses the data to look at the use, abilities and impact of existing AI technology in an area of conflict, while the second level uncovers the theoretical and normative strengths and weaknesses of its use, under the critical theory umbrella. In doing so, the research benefits both those who wish to understand the potential and risks of using such a system as well as those who wish to advance the theory of CTS and promote less harmful counter-terrorism measures.

Chapter 5 - The Case Study: Israel's Use of Counter-Terrorism Predictive AI in the OPT - the Narrative, according to the Israeli Security Forces

In this chapter, I present the ISA's stance on the scope of the AI activity, the historical and legal background to it, the events that led to the creation of the AI, as well as its activity and impact. As explained in the methodological chapter, the first part of the case study analysis reviews the efficacy of the AI tool and explores the ISA's claims about its success using, insofar as possible, similar data to the data the Israeli security forces had. To achieve this, the data in this chapter will be based, as much as possible, on that gathered by the ISA or other Israeli security forces, including publications by senior Israeli officers and interviews with them.

The Historical and Legal Background to the Case Study - The Occupied Palestinian Territories and the ISA

Like many other issues around the Israeli/Palestinian conflict, the question of where and when it started is a highly emotional and political question in the eyes of both sides. For example, whilst the Palestinian narrative points to the beginning of Zionism as a colonial act that is the source of the conflict, the Jewish Israeli narrative is based on the religious cultural concept of the return to the biblical homeland and the unjustified violent Palestinian reaction to it as being the start of the conflict (Dajani Daoudi and Barakat 2013). The involvement of the ISA (under various names) in the conflict can be traced to the early 1940s as part of the 'Haganah', the main Jewish underground movement (Perri 1999). However, as this research is very much focused on a specific time and location, as well as on specific events, the political/legal background to it could be narrowed and could start with the 1967 war and the territories that were occupied by Israel

during it and which are controlled by it to this day. This choice is logical since the case study, as explained above, mostly refers to the 1967 Occupied Palestinian Territories in the West Bank and its Palestinian residents.[12]

In the more than 50 years since the Israeli occupation of the OPT, the official Israeli legal authorities, using the term in international law of 'belligerent occupation',[13] have created what Ben Naftali has identified as 'possibly the most legalized such regime in world history' (Ben-Naftali et al. 2019).[14] This elaborate legal regime was created first, as a tool for the military to control the occupied population (Shehadeh 1988; Hajjar 2005), but was later used for other purposes such as obtaining land, creating and developing settlements (Kretzmer 2012), and even managing peace accord agreements (Shehadeh 1997). Having said that, one aspect of Israel's control of the occupied territories has remained under legalised secrecy and above intervention and that is the jurisdiction, activities and abilities of the ISA in the OPT.

[12] The Occupied Palestinian Territories (OPT) are called, by most Israelis, the 'Occupied Territories' or the 'Held Territories' or just 'the territories', all as part of the Israeli right-wing narrative of 'there are no Palestinian people' (Kampf 2012). However, in order to not take from the rights of Palestinians in this research also, I will use the term Occupied Palestinian Territories (OPT) when referring to the area relevant to the research. The OPT has been changed into many shapes and forms during the years under Israeli occupation, as different regimes have controlled the areas that were captured in the 1967 war (the Sinai Peninsula, the Gaza Strip, the West Bank and East Jerusalem, and the Golan Heights). But for the purposes of this research, since the AI tool has been activated only in the West Bank and East Jerusalem, I will therefore refer to the OPT when I refer to that part of the land.

[13] For more about how international law defines belligerent occupation, and especially in relation to Israel/Palestine, see: R. St. J. Macdonald and Gerhard Von Glahn, 'The Occupation Of Enemy Territory: A Commentary On The Law And Practice Of Belligerent Occupation' (1961) 27 The Canadian Journal of Economics and Political Science 113, Yutaka Arai-Takahashi, 'Law-Making And The Judicial Guarantees In Occupied Territories', *The 1949 Geneva Conventions - A Commentary* (Oxford University Press 2015) 1421. Yoram Dinstein, *The International Law Of Belligerent Occupation* (Cambridge University Press 2009) 108-116.

[14] Ben Naftali recognises several reasons for this. First, the fact that the mission to plan what an Israeli occupation looks like was given to the army legal department. Secondly, all of the military commanders of the occupied territories loved their legislative powers. And thirdly, the Israeli Supreme Court has opened its doors to Palestinian petitions against the military.

The ISA operated based on Israeli government residual power until the late 1990s without any specific law to give it authority.[15] Following the Israeli Supreme Court decision regarding the use of torture[16] as a method of investigation, the ISA law[17] was enacted and defined the organisation's goals and structure (Mann and Shatz 2010). Part of the law is dedicated to the ISA's authority to collect and obtain digital information. This authority is wide-ranging and as Bachar, who was the ISA's legal counsel, identifies:

> 'The existing legislative arrangement in the ISA Law established a dual system, unusual in its intensity, of a database of communications that is physically stored at the ISA as well as a regime for the use of this database that is entirely under the authority of the head of the ISA. The existence of such a large-scale database of all citizens and of powerful technological analysis tools capable of segmenting the data and bringing about a very intrusive real-life monitoring capability, makes it possible to maintain large-scale and intrusive "constant monitoring" of Israeli citizens and residents' (Bachar 2020).

The ISA law still maintains a chain of approval for accessing and collecting data on citizens and residents.[18] Having said that, the ISA law, as an Israeli law, defines the ISA's powers inside Israel, and it does not define its powers in the OPT. There, the ISA powers to operate are subject to the powers given to it by the military commanders of the OPT over the years. Like most military

[15] Historically, the ISA and Israeli Mossad are affiliated to the office of the Israeli Prime Minister, as David Ben Gurion, the first Prime Minister of Israel, wanted control over them for his fight against Jewish right-wing terror organisations, and no other Prime Minister wanted to relinquish this power to the Ministry of Defence.

[16] In HCJ 5100/94 ILDC 2115 (IL 1999), the Israeli Supreme Court ruled that torture is an illegal method of investigation (although it does continue to say that, in extreme cases, a defence of necessity might be given to investigators who use it). The court also detailed the problems in applying special investigation methods without clear jurisdiction in law.

[17] The ISA Law (2002).

[18] Article 11 of the ISA Law (2002).

legislation, the powers that were given are wide, comfortably vague, and include a general authorisation to operate with all the powers a soldier and a member of the police possess in the OPT.[19] As the OPT are, according to the Israeli authorities, held in 'belligerent occupation', the powers given to the ISA include, among others, the power to use and pay agents, to detain, arrest, investigate, enter and search any premises without a warrant and to obtain any digital information about any resident gathered by local cellular and internet providers without the need for pre-judicial approval and to keep this data indefinitely (Bachar 2020).

This almost limitless power has led to the ISA becoming one of the most influential bodies in the daily lives of Palestinians since, apart from its investigatory powers, the ISA is also involved in the daily lives of Palestinians through the permit regime which controls the movement of Palestinians inside and outside of the OPT and therefore defines, among other things, their academic futures and careers (Berda 2017a). It is important to mention that the ISA is, relatively speaking, not a big organisation (approx. 7,000 employees) and it is the constant presence of the Israeli Army in the OPT that facilitates the effective control of the ISA (Bachar 2020). In terms of intelligence gathering, the ISA is tightly coordinated and shares resources with the Israeli army intelligence branch - 'AMAN' - which has very sophisticated data gathering technology (called 'Unit 8200') (Goffman 2019).

The ISA and AMAN's complete technological intelligence control over the OPT developed through the years as more and more technology entered the area. SIGINT collected data was regarded, until very recently, as only an addition to the ISA's original HUMINT agents-based data (Barbing

[19] Article 6 of the Order Regarding Security Provisions (2009).

and Glick 2019). According to Ronen, the OPT was divided into areas and sub areas of control for ISA operatives just as the 1967 war ended and they were instructed to recruit informers from the local population in order to gather intelligence (Ronen 1989). Over the years, and especially during and after the First Intifada, and as Fatah and later Hamas began operating in the OPT, the ISA became heavily dependent on its field agents and their informers (Perri 1999) which, as Hofnung (2017) suggests, might have reached the scale of tens of thousands at some point. The information from those agents was supplemented by the ISA's investigation unit. The investigation powers of the ISA in the OPT are defined mainly by Israeli military legislation and the emergency regulations of the British mandate.[20] Those powers allow ISA instigators to incorporate a variety of official and unofficial investigation techniques, which include long pre-indictment arrest periods, a prohibition on access to legal advice, a ban on audio-visual recordings of investigations, lengthy investigations and sleep deprivation, solitary confinement and, in extreme cases, physical torture, all of which helped the ISA to reach a 98% detailed confession rate from their investigations (Ramati and Torn Hibler 2021). The ISA modus operandi prior to the case study period was therefore mainly based on its extensive network of informers with the support of its investigation units and the digital information collection unit (Pedahzur 2009; Numa and Liraz 2019).

[20] In 1945, the British Mandate government in Palestine enacted the Defence (Emergency) Regulations. They included, in part, provisions against illegal immigration, establishing military tribunals to try civilians without granting the right of appeal, allowing sweeping searches and seizures, prohibiting the publication of books and newspapers, demolishing houses, detaining individuals administratively for an indefinite period, sealing off territories, imposing curfews and declaring illegal organisations. Israel adopted them in 1948, and in 1967 they were adopted to the law of the Occupied Palestinian Territories as part of a military order 'freezing' the legal situation that existed there. Israel argues that the Defence Regulations were part of domestic law in the Occupied Territories prior to occupation.

The Scope of the Case Study: ISA Definitions of Time, Identity and Geography

The specific wave of violence that led to the implementation of the preventive AI tool has been referred to by many names, including the 'Third Intifada', the 'Knife Intifada', and 'the Lone Wolf Intifada'. The Israeli security services named it 'Operation Godel Hashaa' and this is dated to the period between October 2015 and September 2016[21] (Carmeli 2019).

The AI system was designed and planned to identify only potential Palestinian individual attackers (Barbing and Glick 2019). As a 'Palestinian' is a definition that can include Israeli citizens with a Palestinian identity or Palestinian refugees living outside of Israel/Palestine, a Palestinian, according to the ISA definition for the AI system, was 'an Arabic-speaking person, who resides in the territory of Israeli/Palestine, excluding the Gaza strip, and who is not an Israeli citizen'.[22] This oddly-phrased definition leads to a few sub-definitions: 1) We did not follow, for the purposes of this tool, Palestinians with Israeli citizenship. 2) We did not follow foreign citizens if they do not speak Arabic, so foreign nationals from Arab countries, or from Arab origins were followed. 3) We considered Jerusalem as Israel but, because many Palestinians without Israeli citizenship live in the Jerusalem municipal area, we were also extending the search to those people.

The AI was designed to prevent a specific kind of attack and was not intended to replace all the ISA's previous CT tools. The main purpose of the AI was to pre-identify what the ISA has defined as 'individual attackers' and to create a specific profile for them[23] (Barbing and Glick 2019).

[21] The choice of defining the events of the operation using this specific time period is, as will be presented later, a bit unclear when examining the different variables in the data collected by the ISA about the violence. As explained in the methodology chapter, the time period that will be examined in this research includes six months before and after the defined operation time.

[22] From an interview with retired ISA agent 'DB' on 26 April 2022.

[23] More about the ISA profile of potential attackers can be found later in this chapter.

According to 'BC', the system focused on 'around 200,000 potential attackers'[24] who fitted the initial model.

The geographical scope of the AI activity is derived from the identity definitions. According to Goffman, the surveillance was centred on the West Bank and East Jerusalem (although some of the attacks took place in Israel and some of the attackers were Palestinians with Israeli citizenship) because of the restrictions put in place by the ISA legal team (Goffman 2019).

In conclusion, the scope under which the AI was operated, according to the ISA, was very much defined to a narrow and specific time, specific acts and specific demographic and geographical margins.

Chronology of the Wave of Violence and Creation of the AI

To understand what marked the beginning and the end of the 2015-2016 wave of violence, how those events were perceived by the ISA, and what the ISA's response was, I used, as mentioned in the methodological chapter, several sources. Firstly, I used the ISA's monthly, yearly and special reports. Secondly, I used the Intelligence and Terrorism Information Centre's (ITIC) weekly and special reports.[25] Thirdly, I used published articles and interviews with ISA and AMAN personnel that took place around this time. And finally, I used data gathered from my interviews with ISA personnel. Regarding the number of attacks, as the numbers supplied by the ISA and

[24] From an interview with retired ISA agent 'BC' on 28 April 2022.

[25] The ITIC is a part of the Israel Intelligence Heritage & Commemoration Center (IICC) whose members are veterans from all branches of Israel´s intelligence community - the Institute for Intelligence and Special Operations (Mossad), the Israel Security Agency (ISA), and Israel Defence Intelligence (IDI) - and according to DB and BC, is highly aligned with the ISA.

ITIC are mostly aligned,[26] I will use the statistics from the ISA report as a base. The reports are divided by location and by events. Geographically, the reports are divided into events that took place in Israel, events that took place or were initiated in the Gaza Strip, and events that took place in the OPT and Jerusalem. As for the events themselves, for those in the OPT and Jerusalem, they are divided into five categories: explosive devices (usually improvised pipe bombs), shootings, stabbings, running over with a vehicle, and Molotov throwing.[27]

March 2015 to October 2015 - the Period Prior to the Start of the Wave of Violence, According to the ISA

The ISA's data on attacks when it comes to the months prior to the start of the wave of violence, shows no clear trend apart from sharp fluctuations in Molotov throwing which impacts the overall number of attacks (see Figure 1). According to DB, fluctuations such as this had been common since the 2014 war as, 'in one weekend of demonstrations, there could be 100 Molotov throws across the OPT, so the rise from July to September was not an indicator of things to come'.[28]

[26] ITIC reports count some stone throwing events as terror attacks whilst the ISA does not count stone throwing at all.
[27] All of the categories also include attempts to perform these actions, which means, for example, that a person caught at a checkpoint with a knife will be counted as a stabbing attack in the ISA statistics.
[28] From an interview with retired ISA agent 'DB' on 26 April 2022.

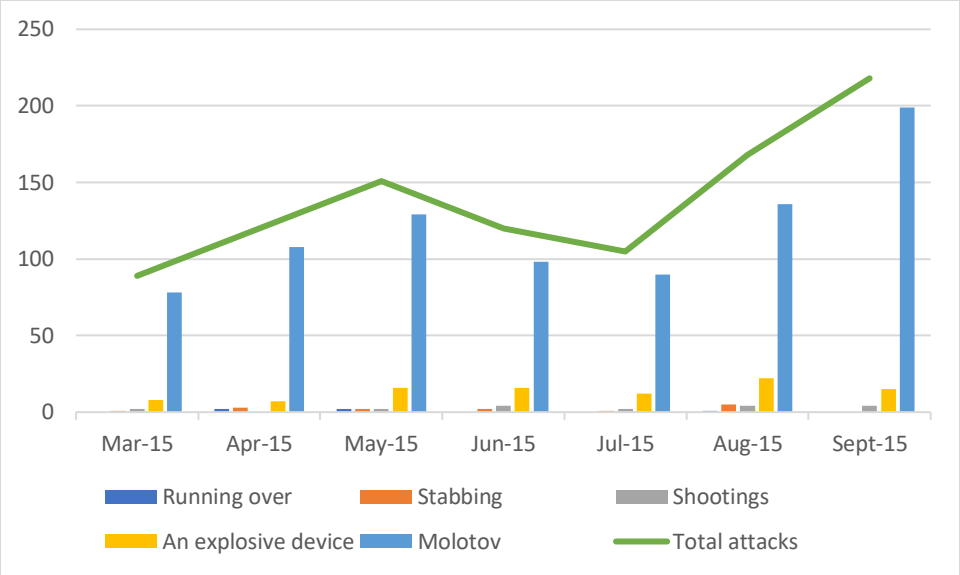


Figure 1. Number of attacks by category March to September 2015

When looking only at the numbers regarding the kind of attacks that led to the creation of the AI tool, which means unplanned individual attacks such as stabbing, running over and shooting, the numbers are relatively similar (between 1 and 5 attacks each month).

The ISA reports from this period do not contain any special remarks about the situation and just mention in the overview a rise or decline in the number of attacks and the people who were injured or killed by them. One different report worth mentioning is that of July 2015, which contains this remark in the overview:

> 'This month, one attack stood out (July 31) - where a Palestinian house in the village of Duma in Samaria was set on fire. Two people were killed in the attack: a baby who was burned to death, and his father who died of his wounds, and two other family members were also injured.'[29]

[29] The ISA July 2015 monthly report from the author's private collection.

This remark is unique in a couple of aspects. Firstly, the use of the phrase 'stood out' which is rarely used in the reports and, secondly, although the identity of the attackers is not mentioned in the report, that specific attack was carried out by Israeli settlers,[30] and data about Israeli Jewish attacks against Palestinians is not presented in the monthly reports.

The ITIC reports for this period were generally more comprehensive and describe in detail any attack or attempted attack. They also attribute an origin to them. During this period, the reports mainly assigned responsibility for attacks carried out by single attackers to big Palestinian organisations. For example, this was the report following a 'running over' attack in May 2015:

> 'A vehicular attack was carried out near the entrance to the village of Alon Shvut in Gush Etzion. Four Israelis were injured, one seriously. The perpetrator had been released from an Israeli jail half a year ago after serving a prison term for throwing stones. This past year there was an increase in the number of vehicular attacks in Jerusalem, Judea and Samaria. Such attacks are supported by Fatah, the Palestinian Authority (PA) does not condemn them and they are considered part of the so-called "popular resistance".'[31]

The rest of the reports for that period described the weekly increase or decrease in violence while the last report prior to the start of the wave of violence identified an increase in violence around Al-Aqsa and Jerusalem, but it did not include any specific identification of a trend or sign when it comes to large-scale events. In general, it seemed like none of the Israeli intelligence bodies

[30] 'West Bank Arson: Dead Palestinian child's father dies of wounds' (2015) BBC. Available at: https://www.bbc.com/news/world-middle-east-33833400

[31] ITIC report May 13-19.

identified any higher than usual danger during that period, as an indication that a new wave of violence was about to erupt.

The ISA Post-Mortem Position on the Conditions that Led to the Beginning of the 2015 Violence and the Individual Attacker Phenomenon

The Israeli government's official position is that the events that started in October 2015, which were later named by the media as the Knife Intifada and by the Army as 'Operation Godel Hashaa', were as a result of an intensive and organised Palestinian incitement campaign (Sikimic 2015). The ISA and the Israeli army's post-mortem analysis took a more complex position regarding the sources of the new wave of violence and the rise of the individual attacker phenomenon. Barbing, who held a senior position at the ISA cyber unit during the events, describes how the ISA identified seven different elements that had a potential influence on the outbreak of violence. It is important to understand how the ISA saw the reasons for the eruption of violence and the change in patterns of the attackers, in order to later evaluate the solutions offered by it. Therefore, in the following paragraphs, I present a short summary of the ISA's analysis of those reasons as presented by Barbing (Barbing and Glick 2019, 130-38).

Al-Aqsa and the Gaza Operation

Israeli actions when it came to the Al-Aqsa mosque, such as the increasing number of attempts by Israeli Jews to pray on the mountain and the attempt to install metal detectors at the entrance to the mosque, were perceived as an attack on the mosque. Al-Aqsa plays a key role in shaping Palestinian identity. Its importance as a Muslim symbol - the third most important mosque in Islam - provides the Palestinian public meaning, as a community in charge of one of the holy places of their religion. This is how the community distinguishes Palestinians from the Muslim

communities of Egypt, Lebanon and Syria. This has led to the Palestinians, and especially the East Jerusalem ones, gaining the role of 'Al-Aqsa protectors'.

The 2014 war in Gaza and its outcome, were also a part of the growing tension in the OPT. According to Barbing, the staggering number of uninvolved Palestinian deaths, especially women and children, and the destruction and suffering the war brought to Gaza created a strong emotional response and evoked continuous calls for revenge.

The Continuous Presence of Israeli Military Forces in Palestinian Cities

Since Operation 'Defensive Shield' in 2002, Palestinians in the OPT have been born into a reality of a constant Israeli military presence, as army forces entered the major Palestinian cities almost every night to arrest suspects, and moving around inside the OPT includes, in most cases, at least one Israeli checkpoint. This reality has played a major part in the feeling of despair and a lack of hope for young Palestinians.

Economic Pressures

The Palestinian economy was highly dependent on the Israeli economy in terms of commerce and even employment, as 20% of Palestinian earnings come from Palestinians who work inside Israel. The unemployment rate among young people was higher than 50%. This was due both to the weak Palestinian economy and to the difficulties young Palestinians faced in obtaining a work permit in Israel. Obtaining a work permit was subject to security clearance and was granted mostly (about 90%) to married people, and the minimum age to obtain one was 24. Although there are many academic institutions in the West Bank where tens of thousands of Palestinians

study, education did not translate into employment, as most of the graduates were unemployed or employed in lower positions that are not relevant to their education.

Political Despair

The Palestinian Authority was considered by many Palestinians as a corrupt body and not a representative one, as no elections to the Palestinian parliament and presidency have been held since 2006. The Palestinian President was also widely criticised for ignoring the corruption of his associates, and 65% of the public believed he should resign. The major organisations were perceived as being detached from the people, and the young especially. The failure of negotiations and the frequent rounds of violence created suspicion and distrust around the possibility for peace and a perception of failure when it comes to the Palestinian leadership. The young people did not identify with the government and accused Israel of preserving it through military coordination. Hamas and other organisations also failed when it came to attracting young Palestinians.

Social Networks

Social networks became available. Information about terrorist attacks, especially in the form of a video or photo, spread within a few minutes to the entire country through a variety of media, and a social network disseminated and created resonance that was unparalleled in traditional media. Despite the lack of a mobile infrastructure, the information consumption capabilities of the new information sources were very high - there were more computers and more telephones among the Palestinian public, especially since 2014. Visual information, in the form of pictures and videos, was the basis for the phenomenon of inspiration or contagion - young Palestinians

sitting on the fence saw the 'successes' of the attacks and draw encouragement from them with a view to carrying out an attack themselves. If Al-Aqsa, the IDF operations in Gaza and personal frustration were the match, social networks were the platform on which the fire flared up and spread rapidly.

Personal Motives

The phenomenon of social networks went hand in hand with the emotional motives of the young terrorists who are not directly related to the conflict or to the tension with IDF soldiers. There was a desire to be like the martyrs of the first and second intifada attacks, who were perceived as Palestinian heroes, and to be displayed across social networks and to gain social recognition.

The emotional motivation was especially pronounced in women who found themselves in a complicated family situation and among young Palestinian men who, for different social reasons, felt inferior to their environment. This phenomenon was seen mainly in villages and less so in cities. The attack was a means of creating an improved social image for young people who were often oppressed, whether because of social status or due to domestic violence. The desire for obtaining honour was directed at Israelis because they were perceived as harming dignity, especially with regard to IDF activity at checkpoints. Thus, the combination of personal connection and desire for revenge produced the threat of the individual attacker.

Organised Incitement

Although incitement from organised terrorism had not been a major rationale for individuals, it was still a possible factor. Hamas and Palestinian Jihad called for an escalation in the violence and expanded their networking activities. Among other things, they tried to initiate campaigns on

social networks, created demonstration videos on how to kill Jews and gave advice on where to act and how. The Palestinian Authority, on the other hand, played the game on both sides - it worked to reduce the incitement phenomenon through fighting against Hamas publications, and at the same time it incited terrorist attacks by glorifying martyrs on Fatah websites and giving money to terrorist families and prisoners. Another factor that served as a space for incitement were the mosques, whose number had increased significantly in recent years. For example, in 2013, there were 1,241 mosques in Judea and Samaria but, by 2014, the number had jumped to 1,892 mosques.

It is important to mention that out of the seven factors that, according to Barbing, the ISA has identified as the reasons for the start of the individual violence phenomenon, the growth in the access and use of social networks was the only new one. All of the other six factors existed at least a decade before, and for some of them, even more.

October 2015 - The Beginning of the Wave of Violence and 'Operation Godel Hashaa'

As already mentioned, the ISA identified the start of the period of individual violence as being October 2015. When examining the ISA data regarding the attacks in this period, in comparison to the previous period, there was an obvious spike in violence in October 2015 compared to September 2015, from 218 to 601 attacks per month, with a drastic change in the results of the attacks, from 2 Palestinian deaths and no Israeli deaths in September to 37 Palestinian deaths and 11 Israeli deaths in October. The sharp rise in deaths comes mainly from the change in the type of attacks, with a sharp rise in attacks carried out by an individual perpetrator. For example,

as seen in *Figure* 2, stabbing attacks went from 0 in September 2015 to 37 in October 2015,[32] which gave this period its popular name, 'The Knife Intifada'.

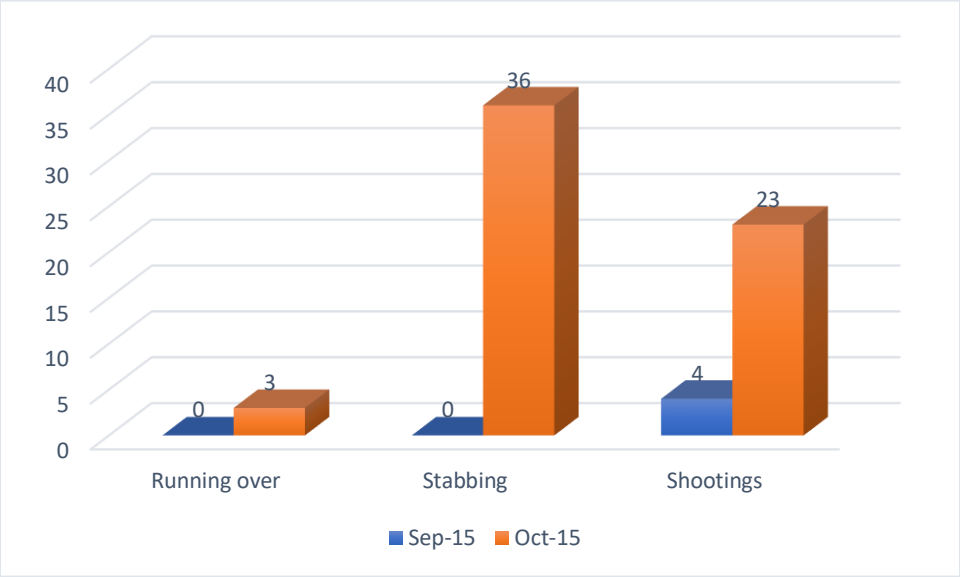


Figure 2. Increase in individual attacks September to October 2015

According to BC, the ISA was not very surprised by the emergence of another wave of violence as 'there were a lot of discussions around Al-Aqsa and the problem of Al-Aqsa is always an explosive one'.[33] The issue that the ISA was not ready for was the type and identity of the attackers. As Barbing identified:

> 'With individual terrorism, there was no clear address that could be investigated. In the past, the agency focused on a number of organisations and terrorist cells that had high signatures, but between the years 2015 and 2017, the agency traced individual threats that did not have clear operational signatures. Moreover, they did not have any intelligence information in the service information systems, as most of them were very

[32] The data is from the ISA October 2015 report.
[33] From an interview with retired ISA agent 'BC' on 28 April 2022.

young and had not taken part in terrorism and violence in the past. This fact made it difficult to detect early and give alerts' (Barbing and Glick 2019).

BC described those first days as 'days of chaos, we were running from one scene to another, trying to understand who those people were and why they were doing what they were doing'.[34] This urgency by the ISA and the other intelligence agencies to find out who the individual attackers were is well documented. On 1 November 2015, a month after the start of the wave, the ISA published a special report titled 'Characteristics of the Current Escalation Wave - October 2015'.[35] The report, which examined 60 attacks in October, focused on understanding who the attackers were and what their motives were, and it identified several joint characteristics: A) A lack of clear ideology or political/organisational affiliation. B) Personal motivation to act based on 'feelings of national, economic, and personal discrimination, including gender (7 women), as well as personal-mental problems'.[36] C) Inspired by social media incitement. D) Influenced by a copycat effect through social media. The report later detailed other statistical information (91% male, 86% single, 82% are between 16 and 24, 67% of attacks were stabbing attacks, 95% of attackers were residents of the OPT and East Jerusalem) but did not infer any conclusions regarding it. A day after the publication of the ISA report, the ITIC published its own report: 'Initial Findings of the Profile of Palestinian Terrorists Who Carried Out Attacks in Israel in the Current Wave of Terrorism'.[37] The report, although it was much more comprehensive than the ISA one and included, for example, a short summary of each attacker's personal information, Facebook

[34] From an interview with retired ISA agent 'DB' on 26 April 2022.
[35] This report is part of the author's collection and was obtained by a freedom of information request.
[36] Ibid.
[37] ITIC special report from 2 November 2015.

pages and pictures, was very similar to the ISA report when it comes to the general characteristics of the attackers. However, one conclusion was unique to it and worth mentioning:

> '…the use of knives has disadvantages and limitations in the current wave of terrorism, as well as advantages. They are not as lethal as guns and suicide bombing attacks, and therefore the number of losses among Israeli civilians and security forces are relatively smaller…. In addition, terrorists who use knives expose themselves to great personal danger and the chances are high that they will be killed while carrying out an attack. Thus, by using knives, Palestinian terrorists undergo great risks to carry out attacks whose fatalities are usually not dramatic… so far there would seem to be a considerable reservoir of young Palestinians driven by various motives to carry out such attacks and maintain the momentum of stabbing attacks in Israel, Judea and Samaria. However, the reservoir may not be bottomless and it is also possible that the wave of stabbing attacks will dwindle and be replaced, at least partially, by a different kind of terrorist attack.'[38]

November 2015 to January 2016 - The Wave Continues

The violence, and especially the individual attacks, continued at a high level over the following four months. When examining the three kinds of attacks the ISA attributes to individual attackers - stabbing, shooting and running over - in this period, in comparison to the period before the start of the violence, the data shows (as seen in *Figure 3*) that, although the attacks did not continue at the same level as October 2015, the level was still much higher than the months before.

[38] Ibid p9.

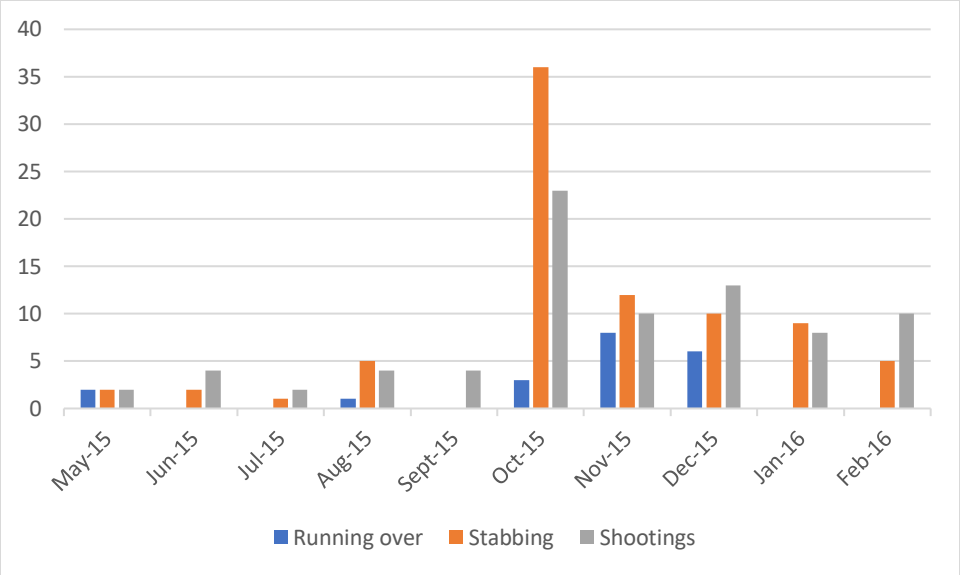


Figure 3. Numbers and types of Individual attacks May 2015 to February 2016

The frustration of the Israeli security forces regarding this pattern and their inability to stop it continued to rise. The army, who at first tried to treat it as just another wave of violence, was not ready. Goffman, who was a senior military commander in the OPT, describes the atmosphere:

> 'That evening, just to myself, I realised: "I am embarrassed" meant I did not understand the problem… the answer was irrelevant and the gap in operational effectiveness was only growing. The pace of operations and the friction and the small changes we made after each debriefing, gave a false assurance that things in the field were moving. But the most embarrassing thing was the fact that even actions that took the initiative and were offensive were left without a logical direction and purpose. I remember the difficult decision of the division commander to stop an offensive operation by three brigades that began following a number of murderous attacks in the Etzion and Judea sectors, because we did not know who to attack' (Goffman 2019, 150).

Similar observations were given by Numa, who was the military commander of the OPT during that time:

> 'The "lawn mowing" mechanism developed during the Second Intifada and the systematic damage to the terrorist infrastructure through intelligence information and night arrests, contributed little to the detection of individual threats in advance. The population protection effort, based on early intelligence from security forces, failed, and the military was forced to "stop the attack on the goal keepers line", according to one of the commanders' (Numa and Liraz 2019, 112).

The army's frustration at the lack of quality intelligence put a lot of pressure on the intelligence agencies and, as DB described it: 'for us, any event or attacker we did not know anything about before the attack was a failure and let's say that, during this period (October 2015 to February 2015, NR), we had mainly failures'.[39] During this period, most of the intelligence efforts were focused on trying to provide the army with tangible targets, which required narrowing the profile of the potential attacker by trying to find connections between the attackers. BC claimed that the ISA was 'all over the place in looking for answers, we were questioning attackers that were alive (not killed during the attack, NR) with the help of a psychologist and going over every detail we had about them to find answers'. An example of these attempts can be found in the ISA report from 16 February entitled 'Exploitation of Family Unification Requests and the Effect of Incitement in the "Palestinian Street" on Terrorists with Personal Problems'.[40] The report suggested, based on three attacks that took place in Jerusalem, three elements that might be

[39] From an interview with retired ISA agent 'DB' on 26 April 2022.
[40] ISA report from the author's private collection.

part of the profile of a potential attacker: one, the use of family unification requests as a tool to move more easily from the OPT to Jerusalem.[41] Two, the personal and psychological problems of attackers. Three, the increasing impact of online incitement. The focus on social media had become more and more prominent in the intelligence discourse. At the end of November 2015, the ITIC published a second report about the profile of the attackers entitled 'Interim Findings of the Profile of Palestinians Who Carried Out Attacks in Judea and Samaria in the Current Terrorist Campaign'.[42] This report, which added to the initial ITIC report new data that had been collected in November, repeated the basic profile findings (lone, male, young, no affiliation, personal reasons) but dedicated a large part of its findings to the influence of social media on the attackers and to their activity on social media prior to the attack. According to the report, which mentioned the word 'Facebook' 112 times, the continuous praising of the dead attackers on social networks had a strong influence on the new attackers who, in turn, became symbols to others. The focus on social media as an important element in the attackers' profiles and activity prior to an attack can be seen also in a series of reports by ITIC, 'Social Networks as a Source of Inspiration and Imitation for Terrorists' (31 January 2016, 02 February 2016, 22 February 2016, 28 February 2016) in which every report analysed the social media activity relating to one of the attacks. For example, the report 'The social networks as a source of inspiration and imitation for terrorists: the case study of two Palestinian youths who carried out a stabbing attack in a supermarket in the commercial area of Sha'ar Benyamin' identified a few elements that are repeated in the social media postings of the attackers prior to their decision to attack: 1) An expressed desire online

[41] Family unification requests are the only bureaucratic tool that can allow Palestinians from the OPT to enter and stay in Israel if they are married to an Israeli citizen or have a first-degree family member that lives in Israel. For more about the family unification of Palestinians, see (Daniel Kasbari 2022).
[42] ITIC special report from 28 November 2015.

that their death will be a martyr death, 2) Facebook friends that reinforce that desire by liking the posts, 3) Adopting previous martyrs as role models, 4) Using Islamic religious terminology, 5) Posting pictures and texts about weapons, 6) Using language that was more adult in nature than the language normally used at that age, 7) No mention of the PA or any other organisation in their feed. This kind of analysis was carried out in all of the reports, which generally repeated and promoted one main idea about the profile of the individual attackers: 'In retrospect, the contents of the posts during the weeks before the attacks reveal the maturing of the terrorists' idea'.[43]

December 2015 to January 2016 - Creating the AI Tool

As already mentioned in the methodological chapter, information about the AI tool does not appear in any official report of the ISA or the ITIC and was not discussed publicly during the time of its implementation. All of the information about it is based on the interviews I have collected, newspaper interviews with agents who were involved in it after the events, and publications in books and journals by the people who were involved in the creation and application of the tool. As the AI tool is presented by the ISA and the army as a great success, there are few people who claim ownership of it, both in the army and in the ISA. However, there is no dispute that it was a joint effort of the ISA and AMAN that led to its creation. Sason Elia, who was the head of the information technology division at the ISA at the time, elaborated in a newspaper interview on the assignment that was given to him at the end of 2015 by the head of the ISA at the time, Nadav Argaman:

[43] ITIC special report from 31 January 2016.

> ‘Argaman wanted something completely different: that our technology would not just be an auxiliary factor, but that it would become one of the most important aspects of the organisation. Take all the technological means that the ISA already has, upgrade them insanely, and bring them all together into one big spy and surveillance system. One that no other country has. Now, go build a “monster” that knows how to collect data from street cameras, SMS, private phone calls, posts on social networks and a thousand and one other sources, saves everything, processes billions of pieces of data a day, and in the end, knows how to bounce a warning to the ISA: Did you hear? Muhammad Hamudi from the village near Tulkarm? He has just taken a gun and gone out towards the checkpoint in order to carry out an attack. Go out and catch him at such and such a point, good luck’ (Shahaf 2020, 4).

Barbing, who was the head of the cyber unit in the ISA and under Elia’s command, describes, in a maybe less colourful way, the ISA’s understanding behind, and decision to develop, the AI tool:

> ‘We understood that the solution to the problem would not be found only in traditional information systems and that warning models based on collecting data through new channels had to be developed. The response included a fundamental change in the collection of information and the development of models for deciphering an event based on the analysis of the characteristics of the attack, the personal characteristics of the attacker, character analysis, learning from previous examples, and more. Instead of looking for an opponent's activity model, we moved on to finding signs of behaviour change. The difference between the two models is that the behaviour was not observed at the physical level - the routine of operations or locating weapons - but the collecting of

> information was required to identify where the perpetrators were digitally - on the network' (Barbing and Glick 2019, 142).

Barbing, BC and DB all referred to the extent of the change the ISA had to go through in order to create this adaptation in its methods. First there was a need to consolidate all the data sources that existed and that required deeper cooperation with other organisations. As DB described it: 'AMAN and the ISA have always co-operated but this thing required us to expose almost everything we had to them and they exposed almost everything they had to us'.[44] According to Barbing, those meetings helped persuade the AMAN technological units - who were focused on collecting data on the main political and terrorist organisations - to widen their collection in terms of population and in terms of the data collected (Barbing and Glick 2019). BC explained the benefits of the co-operation in terms of layers. The ISA had information about houses – 'who is in this house, how many computers they have, what they are doing with them, who they connect to… family? an organisation?'[45] AMAN, which was operating a project 'networking in the Judea and Samaria region', could provide all the geographical knowledge on what was happening outside of the houses, 'who drives which car and to where, who passes which checkpoint and when, who crosses whose path and when, what do they talk about and how are they talking?'[46] To summarise the data sources for the AI tool, the data included: social media information, communication information, movement and location information, visual information, private historical information and personal connection network information. To analyse this amount of data, for all of the Palestinian population all of the time, requires human and technological

[44] From an interview with retired ISA agent 'DB' on 26 April 2022.
[45] From an interview with retired ISA agent 'BC' on 28 April 2022.
[46] Ibid.

capabilities that the ISA and AMAN did not possess. Brigadier General YS, who was the head of the AMAN technological units 8200 at the time, explained in his book how this issue was tackled and how the tool was formed:

> 'Since everyone could be a lone-wolf terrorist, it is impossible to check all of the people all the time….The way to tackle this complicated challenge is with a team consisting of humans and machines, and through the "bounces" and "passes" between this human-machine team. The first step is human. Humans must identify the limits and give examples of characteristics from past lone-wolf terrorists. Then humans and machines together need to formulate the characteristics of potential terrorists, using experience from the past to make predictions in the future. The next step is the prediction that the machine creates through big data about specific suspects. Finally, humans check those suspects and decide how to act' (Y. S. 2021, 76).

According to DB, the first basic profiling was carried out by the ISA just to narrow the number of potential suspects that the AI should monitor: 'At first we went only with these statistics… men, 16-25, presence on social media, no organisational affiliation, high intensity conflict regions… that narrowed it down to around 200,000 potential people'.[47] After that, the AI was given all of the data that was collected from the all of the 2015 previous individual attackers, successful and unsuccessful, 'their social media feeds, their movements prior to the attack and any other pattern we identified'[48] and it created a model to look for, based on those. DB also remembers that they were asked to provide 'some examples of profiles of non-individual attackers to provide the

[47] From interview with retired ISA agent 'DB' on the 26 April 2022.
[48] Ibid.

model with some negative sample'.[49] Once there was a defined audience and a model of a potential attacker, the tool could start working and produce results.

January 2016 to September 2016 - The Activation of the AI Tool

When comparing the number of individual attacks during the second period of 'Operation Godel Hashaa' to the first period (*Figure 4*), it is clear that the number of attacks had been generally further reduced.

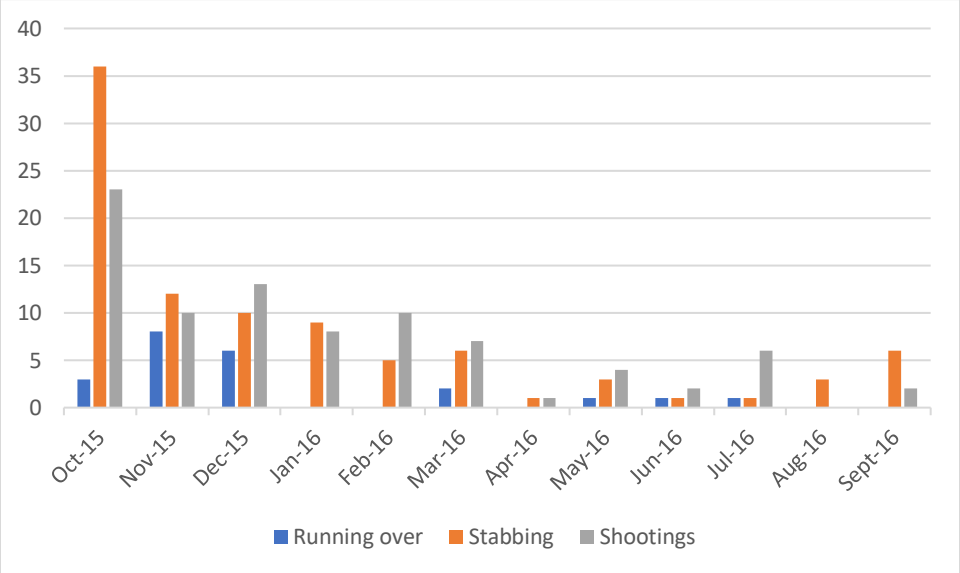


Figure 4. Number of individual attacks October 2015 to September 2016

As mentioned previously, the ISA and ITIC reports from that period do not refer at all to the existence of the AI tool and just refer to the reduction in attacks. However, it is still worth mentioning that, from March 2016, the ISA monthly reports started to include a reference to Jewish attacks on Palestinians. In June 2016, the ITIC published a special report called 'Has the

[49] Ibid.

Wave of Palestinian Terrorism Reached an End?'[50] which concluded that it might ended in April 2016, but that individual attacks would not end.

According to DB and BC, the AI tool officially started working in mid-January, 2016. Barbing described the daily work of the tool in this way:

> 'Stakeholders were monitored according to their activity on the network and were then given a score that stemmed from the level of threat that the stakeholders posed. If, in the past, the perception was binary - belonging to a terrorist organisation or not - now the models included a rating of the threat according to the activity exhibited through social networks and other means. The challenge was to differentiate between a lone threat that is potentially real and expressions of hatred that do not end in activity, "just talk"' (Barbing and Glick 2019, 141).

This ranking system was also described by Elia:

> 'The system knows how to look at the data and say: this person is probably a lone threat, and this person is not. Half a million people have written posts that they want to kill Jews, how do you know who really intends to do that? You use artificial intelligence to produce a score. Above a certain score, it's probably a lone attacker' (Shahaf 2020, 5).

Elia continues and describes what happens after the AI identifies the lone attacker:

> 'This thing produces a very quick closing of the circle between the bodies, including operating the air force, sending a directive to the police, sending my task force. The

[50] ITIC special report from 23 May 2016.

> beauty is that it becomes such a clenched fist. I always admire how a terrorist is shut down with a ten-minute warning' (Shahaf 2020, 5).

This reality of an AI tool that is able to send information and instructions to an operational unit to act, is not corroborated by other sources. According to BC and DB, the AI tool created the ratings and then the ratings were transferred to ISA field agents to assess and decide what should be the course of action. The AI, according to DB, especially at the beginning, was 'shooting all over the place and we were getting dozens of alerts daily, most of them irrelevant'.[51] Support for the fact that the AI did not send direct orders to operating units can be found also in the book by YS (Y. S. 2021, 76) and in an interview with Barbing following his retirement from the ISA:

> 'At one end of this computerised system, there is always an analyst whose job it is to assess how dangerous that same youth is and recommend whether or not he be arrested, called in for questioning, or whether his parents should be called. Everything is examined in accordance with consultations and legal authorisation' (Shragai 2019, 5).

According to BC, the decisions of the agents were fed back into the tool with a ranking according to their relevance and, in that way, 'we continued teaching it all the time what we were really looking for'.[52] DB mentioned that, although the results got better with time in terms of relevance, the number of results did not really change. 'It didn't matter if it was tense or not, (the level of violence, NR) it was still giving us twenty names of people who it identified as being ready for an attack.'[53]

[51] From an interview with retired ISA agent 'DB' on 26 April 2022.
[52] From an interview with retired ISA agent 'BC' on 28 April 2022.
[53] From an interview with retired ISA agent 'DB' on 26 April 2022.

Once the AI tool had created its ranking and the ISA agents had identified the relevant alerts, the next stage was the decision around the preferred way to deal with the threat. DB and BC explain that the relevant alerts included information on the immediate potential to attack, and these cases were treated differently. As DB elaborates:

> 'The system would give us the name and the ranking, mainly based on the last actions this person created on his computer, usually his Facebook page. If we saw someone that fits all the criteria but that also, over the previous hours, had published or carried out an action that could be seen as a decision to act, like posting a picture with a weapon or suddenly shaving his beard and posting a picture of him without it, then there was no question, he had to be detained now. But most of the cases were not like that. In most cases, we identified a potential but no immediate threat'.[54]

According to DB and BC, in those unclear cases, the most common option was to make a phone call. As DB explained, 'you tell him (the suspect, NR) "listen, we know what you are doing and we are watching you". Many times that was enough.'[55] As Barbing explained, those calls were sometimes not enough:

> 'Individuals can be arrested or warned based on their statements or on their online behaviour. Yet deterring an individual is difficult. Therefore the perception of the agency service is that the family and the immediate environment should be deterred, and they will stop the attacker' (Barbing and Glick 2019, 143).

[54] Ibid.
[55] Ibid.

Another option that the ISA used, instead of sending the army an alert or making a phone call, was to alert the Palestinian Authority (PA). According to BC, the PA were, at first, reluctant to cooperate but later 'they (the PA, NR) understood that it also helps them in keeping things under control, so we could call them and tell them this guy needs a talk and they will pick him up and talk with him for a few days'.[56] In some cases, the only choice was to send the army to make an arrest. The suspect would be brought in for questioning and questioned by the ISA and then the Israeli police and, in most cases, they would be transferred to the Israeli military courts.[57] The ISA did not publish any official numbers of potential attacks that were considered as being foiled using the AI tool and many different numbers were thrown around relating to different periods. YS, in his book, says the AI tool has helped to 'prevent tens of lone wolf terror attacks every month' (Y. S. 2021, 76), whilst Argaman told the press in 2017 that, since the start of the events, the AI had helped to stop '1,100 attacks by possible individual terrorists' (Briner 2018). Barbing, in an interview in 2019, estimated that: 'In the past three years – hundreds. With most of them, 50% of the initial intelligence came from the internet and the rest from other intelligence sources the ISA still has' (Shragai 2019). DB explains the difference in numbers, in the way you count a stopped attack:

> 'If you count every one we ever called to do a warning call following an alert, then it could be thousands over the years. If you count people we or the PA arrested, the numbers are much lower. If you count only the people we arrested, the numbers are lower'.[58]

[56] From an interview with retired ISA agent 'BC on 28 April 2022.
[57] More about the legal proceedings will be discussed in the next chapter, which will discuss the legal response.
[58] From an interview with retired ISA agent 'DB' on 26 April 2022.

When asked how effective the AI tool was, the ISA presents a generally united front - that the AI tool had a major influence on the violence both in identifying individual attackers prior to their attack and in reducing the levels of violence in general. For example, the title of the interview with Barbing on the success of the AI tool in the Israeli newspaper Haaretz, is called: 'How Israel Stopped a Third Palestinian Intifada' (Harel 2019). Both BC and DB were a bit more sceptical in their interviews. BC said about the effectiveness of the tool that 'I don't know how many of the people the machine identified would at the end of the day have chosen to attack but it definitely gave us the names of people we could not reach in any other way'.[59] DB had the same notion but added that it was clear that 'all the calls and the arrests have let them know we are watching their Facebook pages, and they can be punished for writing things there, so they became more careful, which I think helped in reducing the violence'.[60]

Conclusion

The narrative set out by the Israeli security agencies, although it varies a little from source to source, is quite cohesive. According to them, several accumulative political and economic conditions led to the wave of violence that started in October 2015. The different agencies did not predict the start of the violence but they were not surprised by it. However, the violence had unique features that had not been seen in previous waves of attacks, especially, when it comes to the phenomenon of unaffiliated individual attackers. A unique AI system was created using expertise and resources from AMAN and the ISA which included the monitoring of social media, phone locations and activities, face recognition cameras, and previously collected intelligence

[59] From an interview with retired ISA agent 'BC' on 28 April 2022.
[60] From an interview with retired ISA agent 'DB' on 26 April 2022.

data, focusing mainly on social media activity as the main indicator. Once the system was activated, it helped in preventing hundreds of attacks and stopped the wave of violence.

The next chapter examines the indictments filed against people who were identified by the AI tool and were arrested by the ISA flowing that identification. Using the indictments, I try to answer a couple of questions. Firstly, can we identify the signals that the AI tool was looking for on the social media of the identified potential attackers? And secondly, how did the Israeli military legal system handle the challenges brought about by the cases generated by the AI tool?

Chapter 6 - The Indictments against Individuals Identified by AI and What Can Be Learned from Them

In the previous chapter, I explored the narrative set out by the Israeli security forces and the ISA regarding the conditions that led to the creation and activation of the ISA's AI tool in a chronological order based mainly on reports, interviews and articles. This chapter will focus on a different kind of data that was collected for this research - the indictments filed with the Israeli military courts against Palestinians identified by the AI tool. During the first part of this chapter, I briefly review the history and structure of the Israeli military law system in the OPT while focusing on the life cycle of a routine case of a person suspected of security offences. The second part is dedicated to an analysis of the indictments themselves. I argue that those indictments, and especially the Facebook posts connected to them, can provide a rare glimpse into what the AI tool was looking for when it scanned social media. In order to find out, I first try to identify which of the indictments collected were based on AI activity, and then use text analysis to find commonalities between the Facebook posts linked to each indictment. The third part examines the legal outcomes of the indictments in order to understand the legal challenges that are connected to a decision to arrest and indict a person based on an identification of an algorithm that this person might be an attacker. To do that, I analyse the court jurisprudence of the AI identified cases and then I discuss their final outcomes.

Israeli Military Law and Courts: History and Operations

As mentioned, the main targets of the AI tool were Palestinians who live in the Occupied Palestinian Territories in the West Bank under the Israeli military legal system. The military courts in the OPT were established based on an Israeli legal position that those territories are not part

of the land of Israel and are held by it as an occupying force (Dinstein 2009). One of the most important duties of the occupier under international humanitarian law is to ensure public order and safety while following the law that was in force prior to the occupation. Article 64(1) of the 4th Geneva Convention details two conditions under which the occupier can change the previous local legislation: if local law is a threat to security, or if it is an obstacle to the application of the Convention. The authority to establish courts can be found under Article 66 of the Convention which lists three requirements for the functioning of military courts: they must be properly constituted, non-political, and located in the occupied territories.

On 7 June 1967, the second day of the Six Days War between Israel and the armies of Egypt, Jordan and Syria, in areas that had already been taken by the Israeli army from neighbouring countries, the Israeli army published and distributed 27-page booklet in Hebrew and Arabic. The first new law written in the booklet was the 'Order Regarding Security Provisions'[61] which established the military courts, defined the ways in which they operate and created a host of new offences. This booklet formed the basis of what would become, over the years, a complex and massive operation that has influenced the lives of thousands of Palestinians in the Occupied Palestinian Territories (Weill 2014). In the fifty plus years that have passed since then, hundreds of military orders (military commander legislation) have been issued by the ever-changing military commanders, covering almost all areas of life in the Occupied Palestinian Territories. A significant part of the legislation was dedicated to the establishment of the military courts and

[61] The 'Order Regarding Security Provisions' has had a few life cycles but the most important ones are, Order 378 that was valid from 1970 to 2009 – Order Regarding Security Provisions; and Order 1651 which is the new codified version that came into being in 2009 and is valid up to today – The Codified Order Regarding Security Provisions.

their duties as well as the creation of new penal provisions mostly through the constant amending of the Order Regarding Security Provisions (Ramati 2019).

The court's geographical jurisdiction has been reduced through the years following a series of agreements and government decisions, first with the return of the Sinai Peninsula to Egypt, the annexation of East Jerusalem and the Golan Heights, and then with the Oslo Accords and the withdrawal from Gaza (see the court current structure in *figure 5*). Although the courts have technical jurisdiction over the actions of the Israeli settlers in the OPT, since the 80s, Israeli settlers have been brought to justice under Israeli law in the Israeli civil courts inside Israel (Ben-Natan 2014). This has created a situation where today the Israeli military courts are active only in the occupied West Bank area, and they only charge Palestinians that are suspected of committing offences there, whilst Israeli citizens who are suspected of committing the same offences, in the same area, are brought to the Israeli civil courts inside Israel and are subjected to Israeli law.

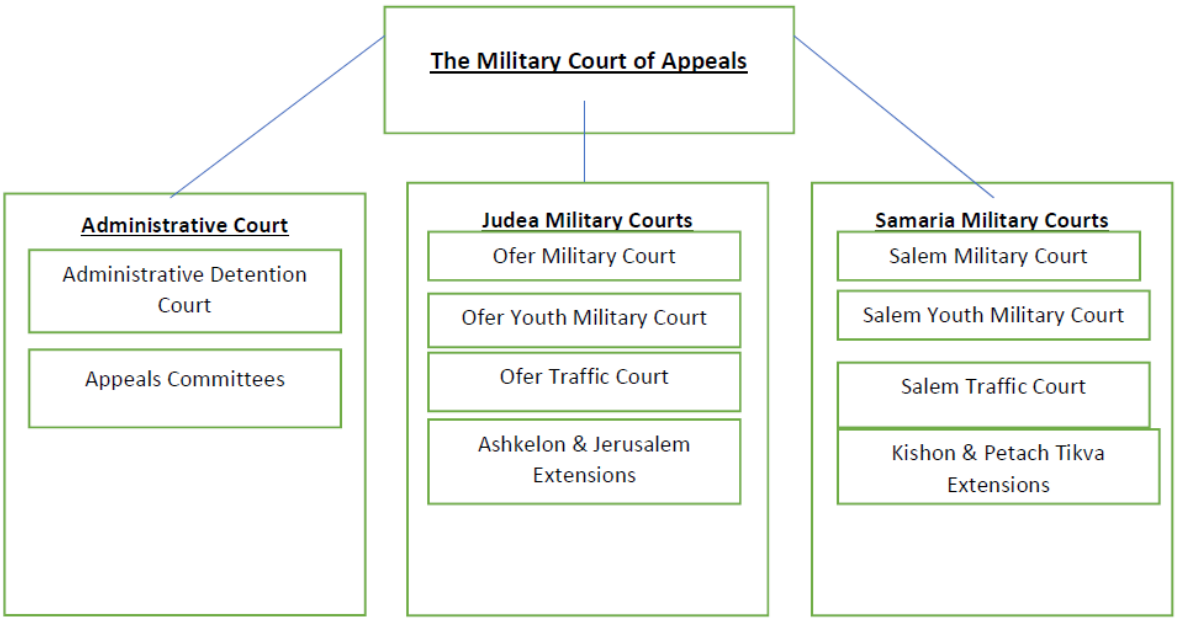


Figure 5. The Israeli military courts structure

The topic of the Israeli military's control of the Occupied Palestinian Territories has been well researched (Dinstein 2009; Gross 2017; Ben-Naftali et al. 2019; Benvenisti 1993). Analyses of the activities and the legality of the Israeli military courts have been conducted by three main groups: military officials, NGOs and academics. Military officials, naturally, present the military court system in a very narrow Israeli perspective (Shamgar 1982; Strashnov 1994; Benisho 2004; Lekach 2017). Palestinian and Israeli NGOs like Addameer, Al Hak, B'Tselem and Yesh Din have been monitoring and documenting the activities of the military courts since the First Intifada and have published detailed reports on several aspects of the courts' activities (Golan 1990; Yavne 2007). These reports provide valuable statistics on the activity of the courts and their impact on the Palestinian population. From a sociological point of view, research on the military courts has been carried out by Lisa Hajjar who explored the courts' activities based on the extensive field work

she conducted. Her research contains valuable data such as interviews with all of the main actors in the courts, from administrators to lawyers, defendants and their families (Hajjar 2005).

Very little attention has been paid to the jurisprudence of the court itself, however. Exceptions are the works of Vitrebo, who examined the military courts' approach in cases with Palestinian minors (Viterbo 2018), Ben-Natan, who was critical of the military courts' tendency to adopt norms from the Israeli legal system (Ben-Natan 2014), and me, looking at references to international law in the courts' rulings (Ramati 2019). Even less research has been carried out regarding law enforcement and prosecution policies and the practices of the police, the ISA and the Israeli military prosecution in the OPT. Hibler and I also looked at the connection between the ISA, the police and the military prosecution when it comes to investigating and prosecuting security offences in the military courts (Ramati and Torn Hibler 2021) and showed the ISA's deep involvement in the criminal proceedings that follow their investigations. ISA involvement in the life of Palestinians in the OPT is not limited to criminal proceedings but there is also the day-to-day bureaucracy tied to the Israeli military occupation (Berda 2017b). However, since until this research, no one had managed to receive a bulk number of indictments against individuals, this research is the first to analyse the systematic use of this legal mechanism, as a tool to answer the ISA's needs in dealing with security challenges.

The Life Cycle of Cases in the Israeli Military Courts

Prior to diving into an analysis of the indictments that were collected during this research, there is a need for a short description of the routine work of the military courts during the period this research is focused on, to understand what those indictments are, and what their purpose is. As

I was a defence lawyer in those courts at the time, this description will be given from my personal perspective. However, all of the statistics that will be mentioned in the following paragraph are from the courts' own statistics and from NGO reports about the courts that were published during those years (Judea and Sumeria 2016; Yavne 2007; B'tselem 2016).

The Israeli military courts' main activity is dealing with what are declared by military law as 'security offences'.[62] These offences yielded between 5,000 and 6,000 cases each year (2015-2016). These security offences can be divided into three clusters - a third are cases of entering Israel without a permit.[63] Another cluster, which is another third of the cases, is dedicated to riot and demonstration offences,[64] and the last cluster and third belongs to offences of hostile terrorist activity.[65] The life cycle of a case in the military courts can be divided into 5 stages:

1) Arrest - Every security offence case in the Israeli military courts starts with an arrest. The suspect is almost never called to come to a police station, and will generally be picked up at night from their house without warning. Another option is the suspect being arrested at a checkpoint or at the riot or demonstration. Any army officer can arrest a Palestinian in the OPT for a period of between 4-8 days before he is taken in front of a judge for the first time.

2) Interrogation - During the 4-8 days of arrest, the suspect is brought to an Israeli police station or an ISA investigative facility for questioning. Questioning can take from a few minutes at the

[62] Another major part of the courts' activity deals with traffic violations, but these will be not discussed here.

[63] This is the most common offence Palestinians are charged with because Israel is the main source of income for Palestinians in the OPT and because of the strict permit regime operated by the ISA (Berda 2017b). The usual punishment for this offence is between 30 and 60 days in prison.

[64] Most of those cases are against young adults and minors who participate in stone throwing against the armed forces and settlements. The common punishment for throwing stones at soldiers is between 6 and 9 months in prison.

[65] This is an umbrella term for all of the more severe security offences, from incitement and membership of an illegal organisation and weapons charges up to killing offences.

police station, if the police consider that there is enough evidence, to two months at an ISA facility if the ISA feels there is more information to be obtained.[66]

3) Indictment - After the police or the ISA have carried out their investigation, they transfer the case file to the Israeli military prosecution which is located at the Israeli military courts. The prosecution reviews the file and decides whether to press charges.[67] The prosecution prepares an indictment, which is where they state all the facts of the case and choose the offence they think is relevant. This indictment is then brought to court together with the case file materials.

4) Remand Bail Hearing - After an indictment is filed, a hearing is held regarding whether the defendant will stay in prison during the trial. The hearing is the first chance for the defendant and their lawyer to view the evidence and understand the charges against them. Generally, the Israeli military prosecution will always oppose bail and the Israeli military courts will generally deny any bail request (in 95% of the cases[68]).

5) Main Case Hearing and Verdict - most of the cases in the military courts (94.5%) end in a plea bargain[69] without any real court hearing, which brings the courts' conviction rate to 99.7% of all cases. The plea bargain usually relates to the punishment and might include some adjustment to the indictment depending on the negotiations between the sides.

[66] The investigations are carried out without the presence of a lawyer and less than 15% of the suspects consult a lawyer prior to the investigation. The rate of confessions obtained during the investigations is very high and stands at 70% for police investigations and 97% for ISA investigations. For more about the difference between ISA and police investigations in the OPT read Ramati and Torn Hibler (2021).

[67] According to a response to the Israeli military courts' freedom of information request filed by Prof. Neta Ziv of Tel Aviv University, the Israeli prosecution decides to press charges in 98% of the cases submitted to it by the police.

[68] Those that are released on bail are involved in cases where it is clear that there is an issue with the evidence, or where it is a clear humanitarian case. In both cases, the defendant will not go back to prison, which makes those hearings the most important ones.

[69] Being arrested without bail and a lack of trust in the courts are the main reasons pushing Palestinian defendants to agree to a plea bargain (Hajjar 2005, 202).

This extremely efficient system creates a result where 94% of all Palestinians in the OPT that are arrested under suspicion of committing a security offence (around 5,000 each year) are convicted and serve prison time as a result.

Prosecuting Palestinians as a Result of Social Media Activity

The 2014 war between Israel and Hamas in the Gaza Strip took place in parallel to an exponential growth in smartphones and social media use (mainly Facebook) in Israel and the OPT (Greenwood 2022). The war was accompanied for the first time by a wave of hateful posts by each side that captured mainstream media attention and brought pressure on the Israeli authorities to react. On 28 August 2014, just two days after the war had ended, a Palestinian man from Hebron was arrested as he was crossing a roadblock and was charged based on his posts during the war. It was the first indictment based on Facebook posts in the Israeli military courts (Kane 2016). This was the first in a wave of indictments that followed, based on social media posts. Reviewing the indictments collected for this research, all the indictments referring to social media activity (Facebook was the only relevant social media network during this period) used the offence of incitement under Article 251 of the Codified Order Regarding Security Provisions (2009). The offence, which is called 'incitement and supporting a hostile organisation' states that:

> 'A person who:
>
> 1) Tries, either verbally or otherwise, to influence public opinion in the region in a manner that may harm public peace or public order, or
>
> 2) Does any act or possesses any object with the intention of making or facilitating an attempt to influence public opinion in the region in a way that might harm public peace or public order, or

3) Publishes words of praise, sympathy or support for a hostile organisation, its actions or its goals, or

4) Commits an act that reveals identification with a hostile organisation with his actions or with his goals or sympathy for them, by waving a flag, by presenting a symbol or slogan or by playing an anthem or slogan, or any similar act that clearly manifests identification or sympathy as mentioned and doing so in a public place or in a way that people who are in a public place can see or hear such a manifestation of sympathy or support, will be sentenced to ten years in prison.'

A hostile organisation is any organisation that has been declared as such by the military commander or any 'person and any group of people whose aim is to harm public security, the IDF forces or the maintenance of public order, in Israel or in an occupied area' according to Article 238 of the Codified Order Regarding Security Provisions (2009).

The list of illegal organisations is constantly updated by the Israeli army and contains over 200 illegal organisations - most of the civil organisations that have been identified by the Israeli army as affiliated, directly or through ideology, with one of the known terror organisations.[70]

This very broad definition of incitement in the military orders has also been interpreted in a wide way by the Israeli military courts and in contrast to the Israeli civil courts (which deals with offences inside Israel, or by Israeli citizens). In the Israeli civil courts, incitement has been interpreted as requiring a probability test checking if the inciting action could create a 'likely/real possibility' that it will impact others (Shinar 2001), whilst the Israeli military courts have ruled

[70] A known example was the decision to declare six well-known Palestinian human rights organisations as hostile organisations, a decision that was challenged by the organisations and supportive western countries. For more, see the B'Tselem declaration from 21 August 2022:
https://www.btselem.org/press_releases/20220821_human_rights_are_not_terrorism

that there is no need to prove such a probability and that it is enough that the text is of an inciting nature regardless of its possible impact.[71]

It should not come as a surprise then that the Israeli military prosecution saw in the felony of incitement a good solution to use for people identified by the ISA's AI tool as potential attackers. The fact that much of the data collected by the AI tool was based on social media activity, together with a wide definition and wide interpretation of an indictment felony, has made it very easy to criminally charge those people.

The Rise of Incitement Charges Following the 2015-2016 Events

The Israeli military courts and the Israeli military prosecution share the same computer system, called 'Magen Tzedek' ('Shield of Justice'). This system has, according to an IDF statement during the freedom of information request proceedings, very basic search abilities. Although indictments can carry several charges with different offences, the system is not able to search for indictments according to a specific offence contained in it. Each indictment is labelled according to what the prosecutor has defined as the main offence. When examining, during the period of the research, the indictments filed by the military prosecution where incitement is labelled as the main offence, the picture is clear.

[71] 'Such behaviour (Incitement NR) should be prohibited even if it does not cause or may not immediately cause the committing of prohibited acts' - Justice Nethanel Beniso, President of the Israeli military courts in 1150/16 Al Haruve v The Military Prosecution.

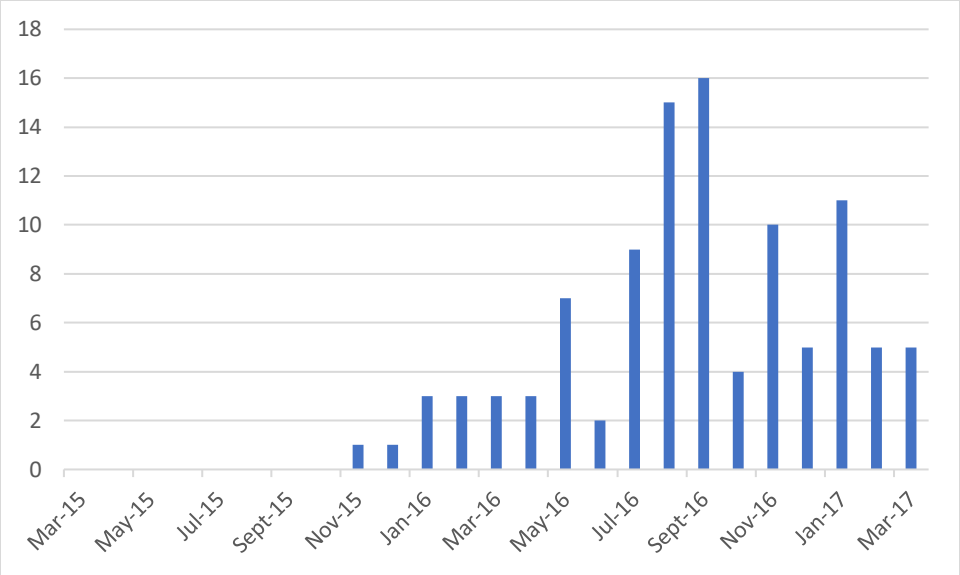


Figure 6. Number of indictments with incitement as main offence from March 2015 to March 2017

Figure 6 very clearly shows that the decision to start using this article followed the rise of individual attacks in October 2015. A similar phenomenon can be observed when examining the number of indictments containing the word 'Facebook' delivered during this period.

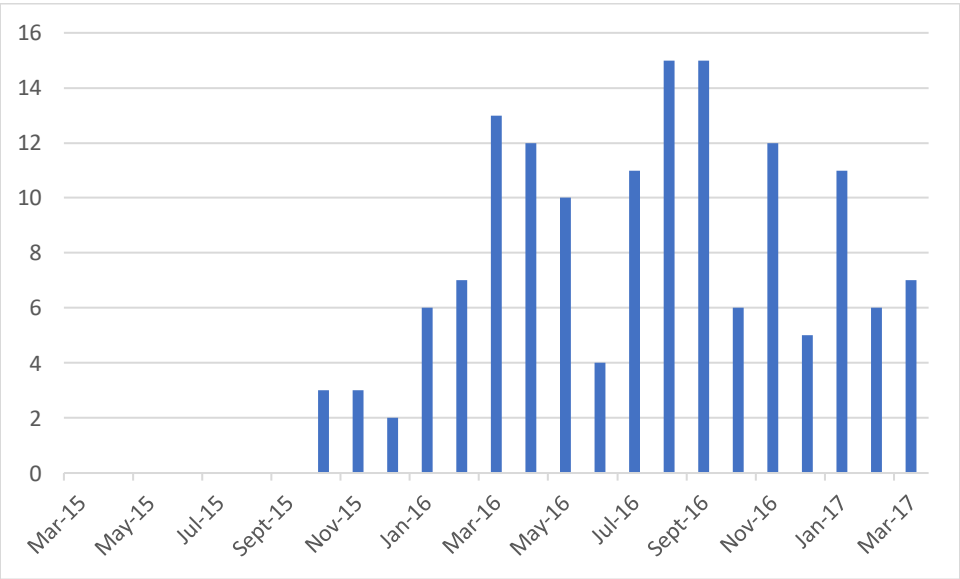


Figure 7. Number of Indictments containing the word 'Facebook' from March 2015 to March 2017

Figure 7 shows that, up to October 2015 - the start date for the wave of violence, according to the ISA - the military prosecution had no interest in what Palestinians were writing on Facebook, the main social platform at that time. The change is abrupt, and very clear, from zero indictments in the six months before the violence started to an average of eight indictments a month for the months that followed.

Identifying AI Indictments from Other Indictments

In the freedom of information request, I asked for all the indictments that contained the words 'Facebook', 'Twitter',[72] 'social media' and all the indictments that mentioned the charge of 'incitement', or any of the 'planning or attempting an attack' charges for the relevant period. After examining the 254 indictments received, I narrowed down the list of possible indictments that could have resulted from the AI tool to 154 indictments. This was done by removing all the indictments that were created prior to the activation of the tool on January 2016, according to BC and DB. The list was further narrowed down by removing all cases that had clear testimonial evidence from others against the defendant prior to their arrest. The presence of evidence such as this indicated that the defendant had been identified by external evidence and not by an AI recommendation. The 154 indictments left were all indictments that contained at least one charge of incitement and had the word 'Facebook' in it.

According to DB and BC, only a portion of the cases identified by AI (the top 5%) led to an arrest and an indictment.[73] On the other hand, as Goffman describes, during the same period, the army command in the OPT had itself been looking for online inciters:

[72] No indictment was found with the word 'Twitter', as it was just starting to gain popularity and the use of it was not common.

[73] From interviews with retired ISA agents 'DB' on 26 April 2022 and 'BC' on 28 April 2022.

> ‘At the beginning of the escalation, when we realised that incitement is one of the most powerful engines to inspire terrorism, we built improvised infrastructures and deployed coordination officers and trackers (Arabic speaking soldiers) to scan sites and look for incitement. It took three months to locate and build an evidentiary infrastructure suitable for an arrest’ (Goffman 2019, 144).

This puts the beginning of both kinds of indictments using the incitement charge on a very similar timeline and reveals that the rise in indictments using incitement as an offence was a result of a combination of the deployment of the ISA AI tool together with the army’s separate attempts to locate the main inciters. So, how can we identify which of the indictments are a result of the AI tool and which are a result of the army’s own proactive work?

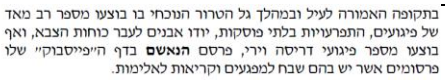
The answer to that lies in the different profiles the army and the ISA tool were looking for. Whilst the army was sending its Arabic-speaking soldiers to look online for people who were using the social network to reach as many people as they could and spread the call for violence, the ISA tool was looking for a very specific combination of indicators based on the profile of a future attacker. This profile, as discussed at length in Chapter five, described troubled young people who are ‘inspired by social media inciters’.[74] This means that the indictments using incitement as an offence, although using the same charge and the same online platform, describe two different kinds of defendants - inciters and followers of incitement.

When looking at the different indictments that refer to social media or use the incitement offence according to this definition, it becomes clearer which indictments are as a result of the army’s proactive attempts and which are from the ISA’s AI tool. In general, indictments by the Israeli

[74] From ‘Characteristics of the Current Escalation Wave - October 2015’ published on 1 November 2015 by the ISA.

military prosecution have a very similar structure to that followed in the AI/army-based incitement indictments as detailed in *Table 2*.

Table 2. The structure of an incitement indictment

1)	Header including the court name, court case number, prosecution case number, police case number and the name of the defendant (redacted for this study).	
2)	A) Name of the offence and its legal reference. B) Details of the offence: starting with 'The defendant in the period (relevant period) has tried…' C) Direct quote from the legal definition of the incitement offence.	
3)	General paragraph referencing the sensitive times: 'During the period mentioned above and during the current wave of terrorism in which dozens of stabbing attacks have been carried out, incessant riots, during which stones were thrown at military forces and even a number of trampling and shooting attacks, the accused posted pictures and	

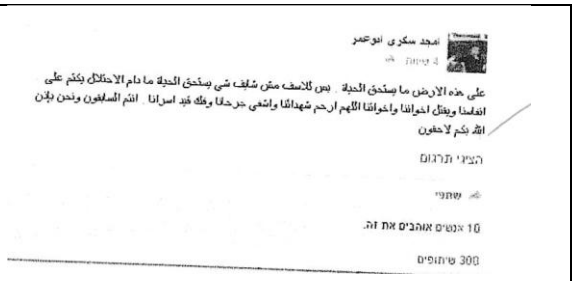
	words of praise for the terrorists on the social network "Facebook".'[75]	
4)	Details of the Facebook page[76] of the defendant and the number of friends/followers: 'During this period, the accused managed a Facebook page named [ ]. This page has about 150 members.'	1. בתקופה זו, ניהל הנאשם עמוד פייסבוק ששמו "[redacted]". לעמוד זה כ-150 חברים.
5)	Details of every relevant Facebook post according to the military prosecution: 'On 28 December 2015, or at a time close to that, the accused updated his background picture which reads 'The martyr writes to me as long as I live and asks me where are you?' 22 people liked this post. On 5 January 2015, or at a time close to it, the accused published a picture of a man riding a horse and holding a Hamas flag and wrote: 'When you see the roof of a flying bus you will know that this attack is for Ayash.	ביום 28/12/2015 או במועד הסמוך לכך, עדכן הנאשם את תמונה הרקע שלו בה כתוב "השהיד מכתב אותי בכל עוד אני חי ושואל אותי איפה אתה?". 22 אנשים אהבו פרסום זה. ביום 05/01/2015 או במועד הסמוך לכך, פרסם הנאשם תמונה של אדם הרוכב על סוס ומחזיק דגל חמאס וכתב "כשאתה רואה גג של אוטובוס עף תדע שהפיגוע הזה לעיאש. זכרו של השהיד מהנדס. הפיגועים של השהידים. השהיד הגיבור, יחיא עיאש ז"ל". 25 אנשים אהבו פרסום זה. ביום 08/01/2016 או במועד הסמוך לכך, פרסם הנאשם פוסט ותוכנו "ולכם במיוחד עם חלש ושפל, יש לכם חיילת שגילה 18 שנים שנכנסת לחברון וסוגרת אותה ולא רואים אפילו גבר אחד מרים את קולו. הם נכנסים ולא מתחשבים בנשים ולא בילדים ואיש אינו יכול לדבר איתם. אתם כמו עדרים. גברים בעיני עצמכם. אתם רוצים לשחרר אותה אבל מה שמגיע לכן זה נעל בפרצוף". 15 אנשים אהבו פרסום זה. ביום 09/01/2016 או במועד הסמוך לכך, פרסם הנאשם תמונה בה רואים אדם וכתוב "השהיד נשאת מלחם". הנאשם כתב "לא תמות לפני שתהיה טהור, השהיד שלנו אחים גרם לעוצר בתל אביב שבוע שלם. עניין שגרם למדינה שלמה, אפילו למזרח התיכון כולו להיות חסר אונים. שאלוהים ירחם עליך גיבור". 16 אנשים אהבו פרסום זה.

[75] Several very similar versions appear in all of the indictments, with slight changes according to the specific period they were written.

[76] The posts were translated into Hebrew on the indictment itself with the original text in Arabic attached to the indictment.

	The memory of the martyr is an engineer. The attacks of the martyrs. The heroic martyr, Yahya Ayash.' 25 people liked this post. On 8 January 2016, or at a time close to it, the defendant published a post with the content: 'And you especially are a weak and lowly people, you have a female soldier who is 18 years old who enters Hebron and closes it and you don't see even one man raise his voice. They enter and do not consider women or children and no one can talk to them. You are like herds. Men in your own eyes. You want to free her, but what you deserve is a shoe in the face.' 15 people liked this post. On 9 January 2016, or at a time close to that, the accused published a picture in which a person is seen and it says 'The Martyr Carrying Bread'. The accused wrote: 'You will not die before, you are pure, our martyr brothers caused a curfew in Tel Aviv for a	

	whole week. A matter that caused an entire country, even the entire Middle East to be helpless. May God have mercy on you, hero.' 16 people liked this post.'	
6	A concluding sentence stating: 'By doing the above, the accused incited and supported a hostile organisation.'	בעשותו את האמור לעיל, הסית ותמך הנאשם בארגון עוין.
7	The witness/evidence lists include only the defendant's testimony taken after the arrest, and copies of his social media posts: 'Prosecution witnesses: 1. Major [ ] - interrogator and translator of the defendant's statement. 2. Photos taken from the defendant's Facebook account.' This is an important indicator that the defendant was identified by the army/ISA, based on his posts online.	**עדי התביעה:** 1. רס"ר [redacted] (גובה ומתרגם אמרת הנאשם). 2. תמונות שהוצאו מחשבון הפייסבוק של הנאשם.

8	A page with photocopies of the printed original posts in Arabic, which are attributed to the defendant's Facebook account.[77]	

The differences between indictments based on the army's proactive search of Palestinian inciters and indictments based on the ISA AI-tool can be found mainly in sections four, five and eight in *Table 2,* which describes the Facebook page of the defendant, their posts and the reactions to the posts. Whilst the army was using its soldiers to go on Facebook and find users with popular public accounts who were spreading calls against Israel or support for attacks and attackers, the AI tool was looking for users who were more private and less popular and where there were alarming signals around recent activity on the platform. Those different user profiles made distinguishing between ISA-AI-origin indictments and army-origin indictments possible. The classification was based on three elements that, combined, led to the decision regarding the source of the indictment.

1) **Facebook Accounts and Details**:

In the indictments against Facebook users who were identified as inciters by the army, the users had an active Facebook page, most of the time with a name that wasn't the user's name but a political name, had 1,000+ friends and had non-friend followers. For example:

[77] This specific post text is from Indictment 1172/16 from the FOI request saying: 'there aren't people who deserve to be loved on this earth, and I don't see anybody worthy of "blood money" for as long as the occupation continues to murder our brothers and sisters. May Allah free our prisoners. RIP our martyrs, you were the first to die and we will join you'.

a) ‘During the aforementioned period, the defendant managed a Facebook page called My Soul for Palestine, the Facebook page had 1,122 members and 194 followers.’[78]

b) ‘The defendant managed a Facebook profile under the name "Zizou Aweda" in the English language. This profile has an unknown number of members and 26,262 followers.’[79]

c) ‘The radio broadcasts of the Al-Sanabel station are filmed and broadcast live on the Facebook page of the Al-Sanabel radio station. This page has 58,917 followers.’[80]

On the other hand, indictments against Facebook users who were flagged by the AI do not describe followers and their Facebook friends did not exceed a few hundred or they are not mentioned at all. For example:

a) ‘During the aforementioned period, the accused managed a Facebook page named “Yosef Farhan Abu Thaar” in Arabic.’[81]

b) ‘During the above-mentioned period, in Selvad or in a place close to it, the accused managed a Facebook page called “Mourad Kwasma” with 149 friends.’[82]

c) ‘The accused manages a page on the social network “Facebook” called “Mohama in Karaforum”. This page has about 300 members.’[83]

2) Number of Publications and Reactions to Them:

Another indication as to the kind of indictment can be found in the number of publications mentioned in the indictment and other users’ reactions and interactions with them.

[78] Indictment 2730/16 from the FOI request.
[79] Indictment 3231/16 from the FOI request.
[80] Indictment 2964/16 from the FOI request.
[81] Indictment 2966/16 from the FOI request.
[82] Indictment 4959/16 from the FOI request.
[83] Indictment 1345/16 from the FOI request.

Indictments against users who are considered by the army to be 'inciters' had a lot more publications mentioned and had a much bigger impact. For example, indictment 2959/16 from 2016 describes 44 different posts uploaded by the user and each of them is dated and translated and they mention the amount of likes they received - from 600 likes, which was the lowest, up to thousands of likes for the most popular post. They also mention the number of times the post was shared, which was at least a dozen for each post.

Indictments against users who were identified by the AI tool tend to be much shorter and the posts mentioned in them less popular. For example, in indictment 1232/16 from January 2016, the whole indictment is based on a single post by the user, a post that was viewed by 59 people and which received 10 likes and was not shared.

3) Content of the Publications and the Last Post Prior to Arrest

The third indication as to the source of the indictment, because of the ISA AI tool or the army, can be found when examining the content of the publications. As explained above, the army-based arrests and indictments usually contained a lot of publications of various types. Because of the number of publications, it was difficult to find a specific trend among them, apart from the fact that they were calling people to action, or just reporting on recent Palestinian attacks or Israeli atrocities. On the other hand, the AI-based indictments had some specific similarities when it came to the posts that appeared in them.

a) Posts that referred to individual actions, feelings or deliberations. Or posts showing or connecting the user with potential weapons. For example,

- A post saying ‘Rest assured the blood of the martyrs was not in vain. This is my love in which I was born and for her I will sacrifice my life, therefore do not be sad because of me, and may I be a martyr in your land’.[84]
- A post showing a picture with the title ‘I’m the next martyr’ with the following text ‘God, let me be a martyr, I miss the martyrs and you, God’.[85]
- A post showing a picture of the body of a dead Palestinian attacker with the text ‘Mother of the martyr, bless you, I wish my mother was in your place’.[86]
- A post showing a picture of a man wearing a cap, holding a weapon with the text ‘Where is the martyr, we are the heroes, we are Abu Amar's men. We love the rifle’.[87]
- A post showing a picture of the person wearing a balaclava and holding a rifle with the text ‘Good evening respectable people – from the heart of the event’.[88]

b) Another good indication was if the last post on the indictment was posted close to the arrest date. According to the description of AI-tool activity, as mentioned in Chapter five, a post on social media can usually trigger an alert that would lead to a quick arrest. Therefore, for many of the AI indictments, there was a short period of time between when the last allegedly inciting post was posted and the time of the arrest. For example:

[84] Indictment 1216/16 from the FOI request.
[85] Indictment 1338/16 from the FOI request.
[86] Indictment 1542/16 from the FOI request.
[87] Indictment 1383/16 from the FOI request.
[88] Indictment 1188/17 from the FOI request.

- 1345/16 - Last post in indictment 17 February 2016 – same as arrest day.[89]
- 1463/16 - Last post in indictment 27 February 2016 – arrested two days later.[90]
- 1622/16 - Last post in indictment 16 March 2016 – arrested the next day.[91]
- 1623/16 - Last post in indictment 21 March 2016 – arrested the next day.[92]

Indictment Classification and Results

Using the differences between the army human-search-based indictments and the ISA AI ones described above, I classified the indictments as belonging to one of the two groups (army indictment or ISA AI indictments) where an indictment clearly had at least two out of the three characteristics for that type of indictment, and no clear characteristics of the other group. For example, if an indictment described a very popular account with many friends and followers, and the indictment mentioned many posts with a strong interaction signal and there was no post relating to a personal thought/action or weapon close to the time of arrest, then it was clearly an army-based indictment that was created as a result of manual labour by military personnel, attempting to identify the leading inciters. On the other hand, if an indictment described an account with no followers, and up to a couple of hundred friends, where the posts received very little interaction or described personal thoughts/actions or weapons and were posted close to the arrest date, then this indictment was put together based on the ISA AI tool. Using this classification, I examined the 154 relevant indictments

[89] Indictment 1345/16 from the FOI request.
[90] Indictment 1463/16 from the FOI request.
[91] Indictment 1622/16 from the FOI request.
[92] Indictment 1623/16 from the FOI request.

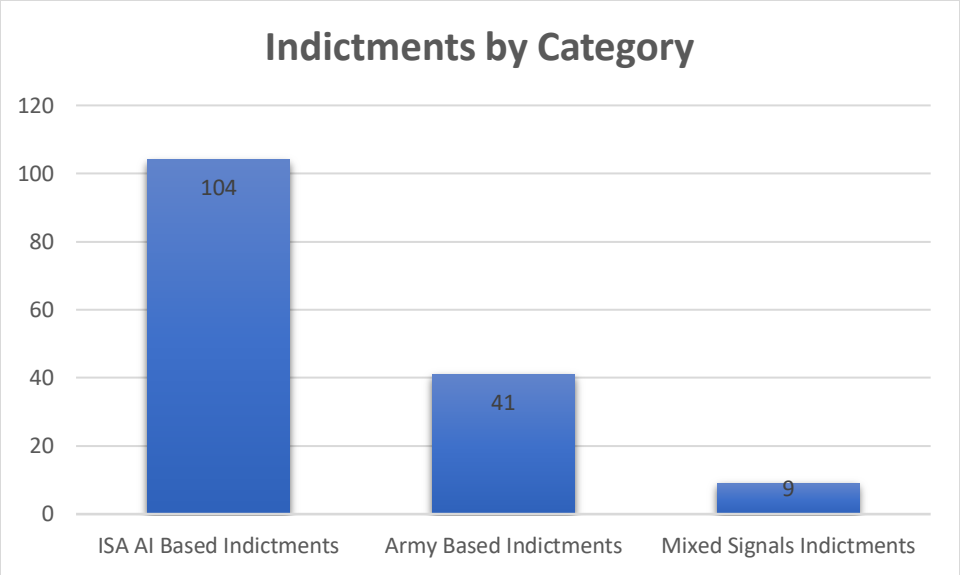


Figure 8. Number of indictments by category

Only 9 (5.8%) of the indictments observed had either no clear signals or mixed signals that made them difficult to allocate to one of the groups. Four of them had 1,000+ friends but very clear content that was both personal and becoming more militant towards the time of the arrest. Five of them had very few interactions but the content of the posts seemed very banal and generic. The rest of the indictments had very clear signals that indicated they belonged to either army-based indictments - 41 (26.6%) - or ISA AI-based indictments - 104 (64%) (see *Figure 8*).

At this point, it is important to raise an obvious caveat - I could not verify my conclusions with an outside source. This occurred because my only direct sources regarding the ISA's AI activity (DB and BC), although involved in some parts of the AI activity, were not directly in charge of submitting every case to the military prosecution, and also because the information could not be obtained through a freedom of information request. Having said that, there are strong indications from secondary sources that support the findings. Firstly, the fact that the army chose an incitement felony for the ISA cases and the profile of those users can be found in the interview with the former ISA cyber unit chief, Erik Barbing (Shragai 2019). Secondly, the number of

indictments fits DB's estimation that, out of 2,000 AI identifications, only 5% were charged.[93] These two sources support the assumption that, even if there is a small margin of error in the classification method, most of the indictments identified are AI-based indictments.

Analysing the AI-Based Indictments

The 104 indictments identified as having been created following ISA AI tool identifications contained several types of information that could be used to learn more about the AI model and the profile of a potential attacker. The following analysis of the indictments starts with identifying some general trends and information that arose from the indictments, and then focuses on an analysis of the posts that were attached to the indictments in an attempt to identify specific signals that could be gleaned from them.

General Signals Arising from the ISA's AI-Based Indictments

From observing the indictments, there are three clear signals that might shed light on the strengths and weaknesses of the AI tool. The first signal is that all the defendants were male, between the ages of 16 and 23. This is not a surprising signal as, according to both Barbing and BC, that was the pool of suspects given to the AI for training data and the base group to look for new attackers. It is not clear from the existing data if the AI was limited to scanning the profiles of only men and only of this age, or if it identified those as factors based on the data that was given to it. However, it should be noted that 7 % of actual attackers during that period were older than 25 and 5% were women,[94] which is not an insignificant number. The lack of any

[93] From interview with retired ISA agent 'DB' on 26 April 2022.
[94] ISA 2016 yearly report, from the author's private collection.

individuals with those characteristics in the indictments might suggest that the ISA decided to limit the search to the main group only - male, between the ages of 16 and 23. The second signal is that, in 18 of the AI-based indictments (17.3%), the defendants were arrested inside the territory of Israel whilst being there without a permit. Although entering Israel to work without a permit is the most common charge against young Palestinians, because a permit is never given before the age of 25, it is interesting to compare this finding to the army-based incitement indictments, where none of the 41 people who were arrested were arrested in Israel. This finding might indicate that the AI considered the geo-location signal of entering Israel as a high-risk variable, which is reasonable, as many of the attacks happened in Israel, especially in Jerusalem, where 14 of the 18 defendants that were arrested in Israel were detained.

Another interesting signal that can be learned from the indictments is the lack of any other signal pointing to a plan to attack. None of the 104 indictments included any specification of any additional info that might support the claim that the people identified were potential attackers. Those signals could have included physical evidence like the locating of a suicide letter or a potential weapon, or testimonies of the defendant or someone else that was aware of the plan to attack. Those kinds of signals are always mentioned in the Israeli military courts' indictments if they are found, and none of the indictments had it. The meaning of this will be discussed in more detail in the next chapter, where I will discuss the possible effectiveness of the tool.

Visual and Textual Analysis of the Posts Identified in the AI-Based Indictments

In order to try and identify some of the signals that the AI tool was looking for on the social media of potential attackers, I compiled a list of all of the posts from the AI-based indictments during a period of one month prior to arrest. This resulted in a list of 614 posts with an average of almost

6 posts per person. The posts can be divided into two groups - visual posts carrying a picture or video - 218 (36%) - and text posts, which were the majority - 396 (64%).

The Visual Posts

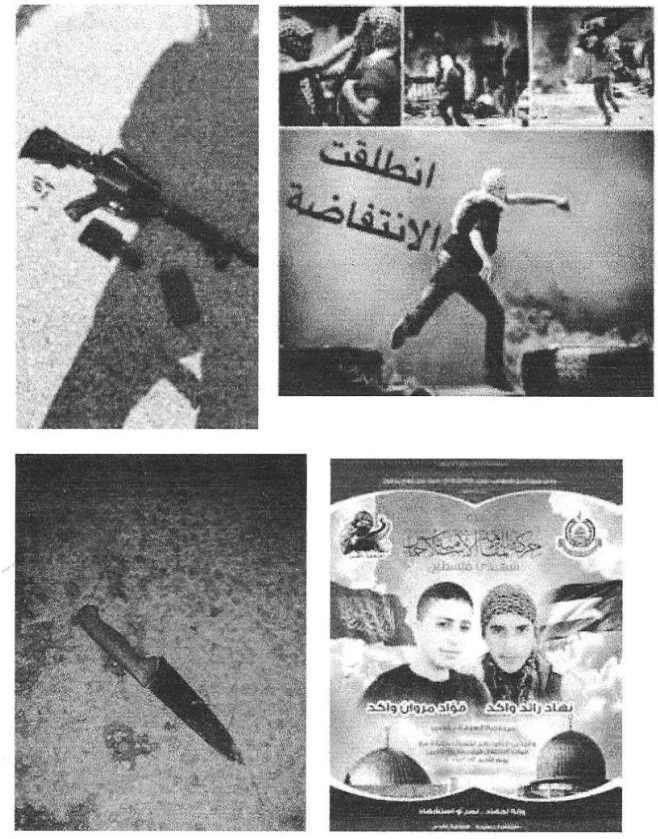


Figure 9. Collection of pictures from the posts attached to the relevant indictments

The visual posts, of which 25% were accompanied by text, were, overall, as can be seen in *Figure 9*, very generic and repetitive. They can be divided into three main groups (see *Figure 10*):

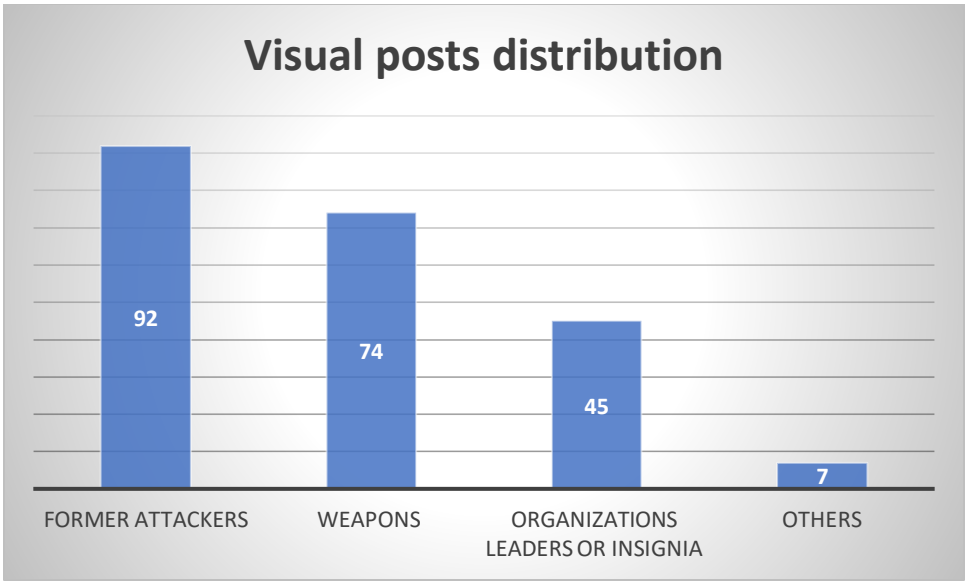


Figure 10. Types of visual posts and their distribution

Former attackers include photos or videos of attackers that had carried out their attack during the months preceding the arrest (78%), or pictures of famous attackers from the period of suicide attacks during the Second Intifada (22%). For example, a picture of Abed al Basat Awda[95] was found four times.

Weapons include pictures or videos of weapons or people holding weapons. The two most popular ones were a picture of a knife with blood on it (37%) and people wearing balaclavas and holding a rifle (22%).

Organisation leaders or insignia include pictures and videos mainly of Hamas insignia and Hamas leaders, with pictures of Sheikh Ahmad Yasin, Hamas founder and religious leader, who was killed in an Israeli attack, being the most popular (38%).

[95] The suicide bomber who blew himself up at the Park Hotel attack and killed 30 Israelis on 27 March 2002.

In conclusion this combination of visual posts in the indictments, did not exposed any specific or unique observation that was carried out by the AI tool, as they all fit the general anti-Israeli Palestinian discourse as identified by the Israeli security forces.[96]

The Text Posts

I referred to any posts that had text in or on it as text posts, including pictures with text attached to them (see *Figure 11*).

Figure 11. Examples of text posts as attached to the indictments[97]

[96] ITIC special report from 23 May 2016.

[97] The top image is from Indictment 1542/16 from the FOI request saying: 'The Heroic martyr (Muhammad El Halabi) may god have mercy on you, hero' accompanied with emojis of a knife. And the bottom image is from Indictment 1216/16 from the FOI request saying: 'I want to leave everything, I don't want to cry, I don't want to complain, I don't want to laugh…. I just want to be gone', as text that appears on a photo.

As explained in Chapter Four, I used an OCR scanning app in Arabic to scan all of the text, and once I had an Excel doc that included each text post separately, I used two Python scripts to analyse the posts, one to look for frequencies and another for thematic connections. The frequency script scanned the text for the most repeated words in all of the posts, and the most repeated combinations of words in each post up to a combination of four words (see *Table 3*).[98]

Table 3. Most repeated words in Facebook posts

Most Repeated Single Words	**Frequency**
Martyr/Martyrs/Martyrdom	305
God	96
Jerusalem	84
Al Aqsa	79
Have Mercy	77
Heaven	70
Attack	65
Homeland	60
Palestine	53

[98] The script ignored a list of 'stop words'. Stop words are general and common words that are usually used for grammatical purposes such as the words 'in', 'or' and 'but' in English. Ignoring these in the search allows more meaningful results.

Knife	48
Occupation	44
Beautiful	39
Mother	35
Hero	27
Muhammad	23
Top 2-Word Combinations Within Posts	
God - Have Mercy	56
Martyr - Hero	33
God - Heaven	26
Attack - Operation	20
Martyrdom - Beautiful	18
Top 4-Word Combinations Within Posts	
Martyr - Hero - God - Have Mercy	15
Hero - Operation - Attack - Knife	13
Al Aqsa - Occupation - Death - Have Mercy	11
Mother - Martyrdom - Cry - Grave	10

The second textual script I used is a semantic similarity script that groups together posts that have semantic similarity into clusters.[99] Each cluster presents one semantic idea which repeats, in some way, in several posts. The three main clusters are identified in *Table 4*.

Table 4. Semantic clusters of Facebook posts

The Martyr cluster : The hero, the beautiful martyr, waits in heaven.	150 posts
The Jerusalem cluster: A fire and a knife for Al Aqsa and Jerusalem, rise up my people!	80 posts
The Mother cluster: Don't be sad Mother, be proud at my grave.	43 posts

Both script results show very similar findings to the basic profiles created by the ISA following the first attacks in October 2015 (Berbing 2019). Most of the posts praised former attackers (martyrs) and it seems that this was the main signal the ISA was looking for. The root word for martyr (شهيد pronounced 'Shahid') which includes the male and female and plural form, and the act of becoming a martyr was the most popular word in the posts appearing in the indictments (three times more than the second most popular word 'God').

In Palestinian society, this word represents a wide range of cases: 'The Shahīd symbolises contrasting connotations: it is both a religious and a secular icon, one that stands for a hero as well as a victim and one understood to denote an active agent as well as a powerless person

[99] The script starts by processing the sample and organising the data into shared clusters based on the principle of semantic similarity between text fragments. For example, semantic similarity will put these two sentences in the same cluster: 1) I have adopted a cat 2) I have a new pet. Then, the script sorts the final results into clusters depending on each cluster's prevalence to the overall data.

whose death was accidental' (Abdul-Dayyem and Ben-Ze'ev 2020, 2). However, the word, in its Arabic pronunciation, has also entered spoken Hebrew with a very specific meaning referring only to suicide attackers. This tension between Israeli Hebrew speakers' understanding of the word, and the Palestinian use of it reached its peak in the court case against the Israeli Palestinian poet Darin Tatur, who was charged with incitement in an Israeli court for using the phrase 'I will go the way of the martyrs (Shahids)' in one of her poems. The three-year-long court case focused on the meaning of the word 'martyrs' in the poem – 'attackers' or 'victims' (Mor 2019). The Palestinian use of the word Shahid, meaning any victim of Israeli force regardless of whether they were an attacker or just an innocent bystander, was of course rising dramatically at the time as, during the 6 months prior to the beginning of the violence, 18 Palestinians were killed by Israelis in comparison to 140 killed in the following six months.

Accordingly, it was not a surprise that the most popular semantic cluster was a cluster of posts praising the martyrs. This find is not surprising because, admiring previous martyrs and the fear of copycat attacks was one of the main signals identified by the ISA as a potential profile. However, looking at the posts themselves, and especially the semantic cluster, does reveal two observations I found interesting. The first observation relates to the word 'beautiful' in describing the martyrs and martyrdom. The 'beautiful martyr' or the 'beautiful martyrdom' was a very popular combination. Posts like 'Cover the martyr with a flag and leave the face visible towards the moon, how beautiful',[100] 'Everyone is looking at you Shahid, how beautiful you are',[101] 'How beautiful is the place where you died, how beautiful is your presence, Moetaz'[102] and 'How

[100] Indictment 1159/16 from the FOI request.
[101] Indictment 3053/16 from the FOI request.
[102] Indictment 3055/16 from the FOI request.

beautiful is the martyrdom for the homeland and how beautiful is the body'[103] were repeated throughout the posts. The use by young Palestinian men of the word 'beautiful' when describing other young men in the martyr context, is something I had not heard about before nor did I find any mention of this in the ISA and army intelligence profiles or in the relevant literature. Therefore, the relatively high frequency of such expressions in the posts could suggest that this signal was identified by the AI tool as being meaningful. The second observation which I found interesting, was the lack of mention of the '72 virgins' who await a martyr, according to popular culture. Although 150 posts mentioned the martyr and heaven together, only one post out of the 614 posts mentioned the virgins at all: 'Tonight the sounds of the bracelets of the virgins will be heard in heaven, because Muhammad ascends to heaven as a bridegroom'.[104] This find contrasts with the ISA and the army profiles of the attackers that gave significant meaning to that myth as being an important incentive for future attackers. It also contrasts with the literature regarding the use of the '72 virgins' myth by Hamas recruiters for suicide bombers during the Second Intifada (Kruglanski et al. 2009). One possible explanation for the lack of reference to this motive in the posts is that this motive is more of a private motive that has been used by recruiters to tempt attackers, but it is less of a public motive to be proud about in social media posts, where national and religious motives can gain more respect. Another possible explanation might be that the myth is much more significant in Israeli eyes than it is in Palestinian eyes, as it allows Israelis to present Palestinian attackers as being driven by non-political motives, and therefore breaks the connection between Israeli control of the Palestinian territories and the attacks. Another

[103] Indictment 3055/16 from the FOI request.
[104] Indictment 3114/16 from the FOI request.

possible explanation is that the AI model did not find this specific kind of expression relevant to the profile created by it.

Another strong thematic indicator stemming from the posts was the repeated use of the word 'mother'. The concept of the mother of the martyr appears as one of the central semantic clusters and it contrasts with the word 'father' which is not mentioned even once in the posts. As identified by Loadenthal, the role of the Palestinian mother in the martyrdom discourse is pivotal, as her sacrifice is considered bigger than that of the martyr, and her role is to overcome this sacrifice by presenting support for the path of the martyred child (Loadenthal 2014, 185). It is no surprise then, that the cluster involving motherhood included the call 'don't cry' as it presents a perfect image of a mother of a martyr being happy that her child has become one.

To Conclude the Analysis of Signals in the Indictments

From analysing the social media posts used in the indictments against people arrested on foot of AI recommendations, it seems like the AI tool gave the ISA a sample of recommendations of people whose online activity generally fitted the ISA's original profile of an attacker - a young Palestinian man, not extremely popular online, with an expressed interest in the conflict and with an emphasis on martyrdom and a specific interest in the beauty of martyrdom and the mothers of the martyrs. The analysis also showed that there are also some indications that the AI considered the presence of a Palestinian posting such posts inside Israel as a danger signal. Another important observation was the lack of any other signals apart from posting those posts, and in some cases being inside Israel, that can support the claim that those people were planning or about to plan an attack. These findings will be further discussed in the next chapter which

discusses the possible effectiveness of the ISA's AI tool. On the next few pages, I will continue to discuss the indictments that were created based on the AI tool identification but this time from the legal perspective.

The Legal Implications of the AI Indictments

As mentioned in Chapter Two, the literature discussing the possibility of using counter terrorism algorithms to identify individual attackers not only questions the effectiveness of such tools but also discusses the legal hurdles that are involved. In this case, creating a system where an AI tool can identify a possible threat in the actions of a Palestinian, based mainly on their social media activity, leading to an arrest and then to an indictment, is one stage of the process. However, convincing a court to jail someone for it is another matter. In the following paragraphs, I will analyse the Israeli military courts' response to the AI indictments and their final outcome.

The Military Courts' Response to the AI-based Indictments

Israeli military court statistics and Israeli and Palestinian NGO reports on the courts present a clear and bleak picture about the Israeli military court proceedings. According to them, the conviction rate is constantly at 99.7%, and requests to keep the defendants under arrest during their trial are approved 95% of the time (Judea and Sumeria 2016; Hunt 1987; 'The Occupation's Fig Leaf: Israel's Military Law Enforcement System as a Whitewash Mechanism | B'Tselem,' n.d.). These high percentages hide another phenomenon which is that the vast majority of the convictions and arrest requests are carried out by way of a plea bargain or an agreement between the military prosecution and the Palestinian defence attorneys, due to the mistrust most Palestinian attorneys have towards the military judicial system and their workload (Neta Ziv 2018, 752). This phenomenon was not much different in the AI-based indictment cases. Out of the 104

indictments identified as being AI based, there were only seven elaborated decisions that presented some jurisprudence, six of those were about bail, and one about sentencing. All of the other 97 cases of AI indictments ended up in a plea bargain agreement that was made whilst the defendant was under arrest without bail.

The fact that the ISA was using predictive AI to identify lone attackers was only exposed to the public in 2018 in a speech by the head of the ISA at a professional conference (Briner 2018). Up to then, there had been no official acknowledgement of the tool and therefore officially, during the period of this research (2015-2017), the courts had no evidence regarding the source that led to the specific Facebook user being brought to justice. All the courts had in terms of evidence were the Facebook pages attributed to the defendant and the defendant's testimony after the arrest. The two legal questions that are raised in literature regarding the problems of using such an AI tool are the problem of the invasion of the right to privacy and the problem of targeting someone before they commit any violative action. The question of the possible violation of the right to privacy was not discussed at all regarding the ISA-based cases, although it was clear that intrusive surveillance methods were used to obtain the evidence. The reason for that is a decision by the Israeli Supreme Court regarding surveillance on phones or electronic devices of Palestinians in the OPT, that the Israeli law that requires a pre-approved order to conduct such surveillance does not apply to the Israeli police in the OPT in regards to surveillance of the non-Israeli population.[105]

[105] In HCJ 4211/91 El Masri Vs Israel 624 (IL 1993), the Israeli Supreme Court ruled that international humanitarian law does not give any special protection to the protected persons of the occupied territories regarding the right to privacy.

As for the question of how to legally charge someone just for being identified as having the potential to commit an attack, the answer that was found by the military prosecution was to use the incitement offence based on their social media activity. Since the profile of a possible attacker, as identified above, was based on a user with few followers and social media interactions, the court, in some cases, found it difficult to accept the military prosecution's very wide definition of the 'incitement and supporting a hostile organisation' offence. Firstly, the appeal court laid out what should be examined when considering bail for such an indictment:

> 'The degree of risk associated with the publications is assessed according to the content of the publications, and the strength of the call to use violence or assist violence; the number of publications and the length of time they were published; the status and extent of the writer's influence on people's opinions; the number of people who were exposed to the publications, and the extent of their support for the publications.'[106]

The Military Appeals Court later made a distinction between posts that encourage violent actions and calls to support 'hostile organisations':

> 'There is also room to make a distinction between incitement to commit violent acts, which as a rule is more dangerous, and identification with or support for a hostile organisation, for which a high level of seriousness will be required to establish a cause of danger... in my opinion, and further to the distinction I made in the previous part of the decision, which referred to the rationale for the various offences grouped together in section 251, unique circumstances are required to seriously establish a reason for dangerousness that excludes an alternative with regard to supporting a hostile

[106] Judge Benisho in arrest appeal 3044/15, The Military Prosecution Vs Abu Salim (published on Nevo.co.il).

organisation. This is in contrast to words that directly incite violence which, I believe, should be used to establish a reason for danger even without the unique circumstances.'[107]

Those decisions have opened the door for bail release in some of the AI-based indictments and have led to the military prosecution presenting to the court an expert opinion by the ISA, which includes the profile of the recent individual attackers as was created by them, in an attempt to persuade the Military Appeal Court of the need to keep the defendants under arrest as they fit the profile. The expert opinion, as we can learn from the court decision, tried to hint to the court that although the charge against those defendants was incitement, they are really dangerous because they are potential attackers.

The court decision in this case is the most elaborate jurisprudence of the court when it comes to those indictments and therefore I have chosen to quote large portions of it. Firstly, the court presents the general expert opinion:

'The prosecution presented an expert opinion by the ISA on the connection between social networks on the Internet and the terrorist attacks of the past months. According to this opinion, out of 232 attacks carried out in 2015, 159 attacks were carried out by 'individual attackers', who were not members of terrorist infrastructures or recognised terrorist organisations. After examining the characteristics of 156 such attacks, it was found that in 115 of them (73%), there was activity by the perpetrator on at least one social network account, and on 73 of the perpetrators' Facebook profiles (70%),

[107] Judge Benisho in arrest appeal 1150/16, The Military Prosecution Vs Al Harub (published on Nevo.co.il).

"abnormal and extreme" statements were found on the part of the perpetrator, including clear intentions to carry out an attack, and parting words.'[108]

This expert opinion, as given to the court, which was written in February 2016, a month after the activation of the AI tool, was the first time the ISA officially exposed the real reasons for the arrests that were made based on the tool's identification. The expert opinion does not claim that individual attackers were influenced by inciters and that is the reason they ask for the arrest of inciters. The expert opinion says attackers are posting 'abnormal and extreme statements', which are what are described in the indictments; therefore, the people arrested are potential attackers. Interestingly, in that decision, the Israeli military court chose not to question the legal problem that arises from this expert opinion, which is why you charge those people with incitement, although this is not the reason you arrested them, but to question the algorithm's effectiveness. At the second stage of the ruling, the court asks whether the evidence in the opinion is strong and clear enough to create a real correlation between the expression and possible attacks:

'In my opinion, the ISA opinion calls for a number of "cautionary notes": it does sound logical, and points to a connection between postings on the Internet and terrorist attacks; but it does not prove an unequivocal causal connection. The fact that a significant percentage of the perpetrators were active on the Internet or writing extreme content on Facebook does not show that this is the main cause of an attack, nor even a distinct risk factor. One must beware of the cognitive bias of a "fake correlation" (the impression that two events are related or will appear together with high frequency, simply because there is an associative relationship between them) and one must guard against the cognitive

[108] Judge Azmon in arrest appeal 1222/16, The Military Prosecution Vs Hamed (published on Nevo.co.il).

bias of “framing”, according to which things seem more serious or more common than they are, just because of the way they are presented (see Hami Ben-Nun, Cognitive Biases and Judicial Decisions, Sha'are Mishpat 5 (5777), pp. 210 and 211). After all, if 70% of the 73% of the attackers (which is a total of 46% of the attackers) wrote “extreme” content on Facebook before going on the attack, this means that more than half of the individual attackers (54%) did not write any extreme statements on Facebook. In the absence of details about the attackers’ other interests and activities, it is not clear what other factors might have caused them to carry out attacks, apart from internet correspondence. According to the expert opinion itself, it seems that the real risk factor for carrying out attacks was not the opinions and slogans posted by their friends on Facebook, but other factors (for example, watching the news on TV and videos on the internet). It is also possible that victims are also affected by feelings of revenge for the death or injury of a relative or friend, messages from family members and teachers, sermons at the mosque or meetings at the university, reading newspapers, family and social problems, and exposure to the callings of Palestinian public figures and opinion leaders.’[109]

In conclusion, the court asks if looking at and putting pressure on young people who express themselves online is the right approach at all:

‘In addition, it must be examined whether the connection between publications with violent and inciteful language and violent events justifies everything that the military prosecution (and Israel) seeks to achieve: first of all, it must be examined whether this connection is real and distinct, and this, as mentioned, has not yet been proven by the

[109] Ibid.

> opinion (just as opinions are still divided whether computer games and violent movies cause violence among teenagers, and whether mobile phones cause cancer); Is the arrest and prosecution of young people who wrote inflammatory 'posts' the right way of combating individual attackers, or is it a matter of 'searching for the coin under the lamp', treating what is easy to treat, which may lead to the neglect of searching for and dealing with other risk factors (perhaps because it is difficult to detect and arrest the main causes of violence)? It should also be examined whether it is correct to limit free expression on Palestinian social networks, which probably also serve as a valve for releasing feelings of frustration and anger, and whether criminal punishment of expressions of anger will lead to the diversion of the heated discourse on social networks to even more secret and dangerous channels.'[110]

In this unique decision, the court indirectly questions the ISA predictive model's effectiveness. In the narrow sense, is this model logical, and can it identify potential attackers? In the wider sense, will arresting people based on social media expressions reduce the phenomenon of individual attacks in society?

The Move from Incitement Cases to Administrative Arrests

The Military Appeal Court rulings in the cases above, and their decision to release on bail some of the defendants who were charged based on the predictions of the ISA AI tool, led to the military prosecution changing their way of doing things in those cases. As BC describes, the ISA pressured the military prosecution to find a solution: 'We understood why the court was finding it difficult to keep those people under arrest, but we were sure that it was necessary and we did

[110] Ibid.

not care how they were arrested.[111] The solution that was found was to use administrative arrests on anyone who was released by the military court of appeals. Since the ruling in the Hamad case, any case that was brought by the ISA, and which was an AI-based indictment, was first filed as a regular incitement case together with a request to remand until the end of the proceedings. Where the military court ruled to release the defendant on bail, an appeal would be presented to the appeal court. Where the appeal court would also choose to release the defendant, a request would be filed to withhold the release to allow the prosecution to issue an administrative arrest warrant.[112] This practice can be seen in this decision:

> 'One must take into account the fact that this is not a large amount of publications with limited circulation, both in terms of those who follow the publications, and in terms of the limited number of responses received to them. Similarly, I did not get the impression that the defendant has a status that might make the publications particularly influential. The appeal is therefore dismissed. The defendant will be eligible for release under the conditions set by the lower court.'
>
> In view of the prosecution's request, I am delaying the implementation of this decision for 72 hours, in order to consider issuing an administrative arrest warrant in the case of the respondent.'[113]

[111] From interview with retired ISA agent 'BC' on 24 April 2022.

[112] Administrative arrests are pre-emptive arrests without trial and based on security reasons. The logic of these arrests is based on the idea that the security services have secret information pointing to the fact that a person might commit an offence in the future, but exposing such information will hurt the source of the information. Based on this evidence, the military commander of the area issues an arrest warrant which is examined by the court when only the court is exposed to the evidentiary basis, and not the person who was arrested or their defence attorney. For more about the Israeli system of administrative arrests see the European parliament policy paper briefing at: https://www.europarl.europa.eu/RegData/etudes/briefing_note/join/2012/491444/EXPO-AFET_SP(2012)491444_EN.pdf

[113] Judge Benisho in arrest appeal 2933/16 - The Military Prosecution Vs Al Abu Absa (published on Nevo.co.il).

The chances of a Palestinian challenging an administrative arrest warrant in the Israeli military legal system is close to zero, as only one such case has succeeded in the last decade (Krebs 2012). This practice of securing the arrest of people identified by the ISA tool through an administrative arrest has led defence attorneys to stop requesting that their clients be released on bail. A court punishment has a clear end date, but an administrative detention can be theoretically prolonged forever. The upshot of this was that only five of the 104 people indicted with ISA-based indictments were released on bail, all of them prior to the court decision on the Hamed case. The rest of the defendants were either remanded and then sentenced to periods between nine and 12 months in prison in a plea bargain or arrested using an administrative detention order for periods of between four and 12 months.

Conclusion

The process described above presents a very effective system. An AI system is designed to identify people who fit a profile that was fed to it. Those people are arrested and, regardless of the position they take in their interrogation, they are charged with incitement. The courts, although not convinced that this is the right legal mechanism, approve the plea bargains or the administrative warrants that send those people for long periods in prison. This closed tautological system was seen by the ISA as a great success. In the next chapter, I examine this system with more critical eyes and challenge it in terms of its effectiveness, legality, and morality.

Chapter 7 - Is it Terrorism? Is it AI? Is it Preventive? Why Do They Want it? Why Does it Matter?

This chapter examines how the ISA's preventive AI case study fits within the current critical literature on terrorism, predictive artificial intelligence, and counter-terrorism. The chapter consist of three parts. Firstly, I will examine the ISA's assumptions regarding waves of violence, as described from the ISA's point of view in Chapter Five. Questions such as what the reasons were for the outbreak of violence will be addressed as well as who was considered an attacker, what was considered an attack and what was considered as AI when creating the tool. In the second part, I identify the different reasons that pushed the ISA to develop and rely on the AI tool, while presenting three explanations that are outside the discussion on effectiveness. In the third part, I examine the Israeli case study to discuss its potential effectiveness in reducing violence in the short and long term and the very long term.

The ISA AI tool: Background, Conditions and Assumptions

Israel/Palestine: Terrorism and Counter-Terrorism

Throughout this research, I have only lightly touched on the main political and theoretical questions around the Israeli/Palestinian conflict and the term 'terrorism' because the focus of this research is on the theoretical questions that arise from using predictive AI in a counter-terrorism environment. However, as the case study deals with the Israeli/Palestinian conflict and since the overarching framework I am using in this research is Critical Terrorism Studies, there is a need to clarify the main terms used in the context of the ISA tool in the Israeli/Palestinian conflict. As discussed at length in Chapter three, understanding how the terms 'terrorism', 'terrorists' and 'counter-terrorism' are used to advance political agendas is essential to any

academic discussion about them. This is, of course, relevant to the use of those terms in the Israeli/Palestinian context. It does not matter if you look at the Israeli/Palestinian conflict through the prism of international law as an occupation or apartheid regime, or through political science eyes as a settler colonialist situation, or both, it is obvious that Israel has used - and is using - the term terrorism to frame any kind of resistance by Palestinians. As Pappé has identified, 'this characterisation is so gripping that it leaves very few chapters in Palestinian history outside the world of "terrorism" and hardly absolves any of the organisations and personalities that made up the Palestinian national movement from the accusation of being terrorists' (Pappé 2009, 128). This approach is not unique to the IL/PS conflict and has been used in other post-colonial conflicts too (McQuade 2020, 6).

As explained before, whilst legally the words 'terrorist' and 'terrorism' are not commonly used in the Israeli legal system (which prefers the term 'security offences'), in any other sense, 'terrorist' and 'terrorism' have become another word to describe Palestinians in general because, in the eyes of the populist right-wing movement, a Hamas military-wing member from Gaza and a Palestinian with Israeli citizenship who is a member of the Israeli parliament are both 'terrorists' (Dorfman 2022).

This situation is not much better in the official Israeli security forces, as similar counter-terrorism strategies are used against youths who, during demonstrations in their villages, throw stones against heavily protected Israeli soldiers, as they are against suicide bombers on their way to attack a café full of civilians in Tel Aviv. Any violence by any Palestinian against any Israeli is 'terrorism' which requires the use of counter-terrorism measures (Ramati and Torn Hibler 2021). In fact, researchers of the conflict have identified time and time again that actions that have been

framed as counter-terrorism efforts by the Israeli authorities, were in fact just another tool of control over the entire Palestinian population and land (Ben-Naftali et al. 2019; Ben-Natan 2014; Berda 2017a; Hever 2018). The clearest example of such an approach can be seen in the establishment of the illegal Israeli settlements in the OPT which were presented - and still are presented by some - as being essential for the security of Israel and an important tool in the fight against terrorism (S. A. Cohen and Klieman 2019). All of this research mentioned above identifies one re-occurring observation regarding the Israeli security forces: there is a continuous focus on developing short-term tactical solutions against any Palestinian resistance with no long-term strategic or political solution (Catignani 2017). Therefore, when examining the ISA AI tool, it needs to be put in that perspective - another measure in a long chain of measures of control implemented against the Palestinian population.

What Started in October 2015? Why? And By Whom?

In Chapter Five, I presented at length the ISA's analysis of the reasons that brought about the beginning of the wave of violence. According to Berbing, there were seven elements that combined to lead to the rise of the individual attacks phenomenon: 1) tensions at the Al-Aqsa mosque and the 2014 Gaza war, 2) the continuous presence of Israeli Military Forces in Palestinian cities, 3) economic pressures, 4) political desperation, 5) social networks, 6) personal motives, and 7) organised incitement (Berbing 2019, 130). This open and detailed discussion on the social, economic and personal reasons, and the in-depth processes taking place in Palestinian society, can be confusing. At first glance, it looks like the ISA devoted extended and honest research efforts to discovering the deep-rooted causes that led to the rise of the phenomenon. Observations claiming that young Palestinians are deeply impacted by the 'Israeli operations in

Gaza and the high death rate of woman and children' or that the sensitivities of Palestinian society when it comes to Jerusalem are supported by distrust of the 'Israeli authorities' action at the Al-Aqsa mosque' or that constant Israeli army activity in the West Bank has 'shaped the perception of the young Palestinian generation towards an occupied nation whose lands have been taken, full of despair and lacking in hope' (Berbing 2019, 131), give the impression that those were written by NGO activists and not ISA agents. After all, one of the main criticisms of classical terrorism studies is that it carries 'the tendency to study terrorism separately from the social movements, state structures, conflicts, history, contexts, and international relations within which it occurs' (Jackson, Smith, et al. 2009, 219), and that extremely detailed and, on the surface of things, brutally honest descriptions of the situation on Palestinian streets by the ISA, fit this description. However, a more careful reading of the text raises a few interesting questions. The first and foremost is 'where is the ISA's responsibility and contribution to the situation?' As described throughout this research, the ISA's role in controlling daily Palestinian life is a pivotal one. It is almost absurd to read the ISA's analysis of the hardships of a young Palestinian man, as it ignores that those hardships exist because of its policies. For example, Baring describes the financial pressures that affect young Palestinians:

> 'The unemployment rate among young people is higher than 50 per cent. This is due both to the weak Palestinian economy and because it is especially difficult for young people to get a work permit in Israel. Obtaining a work permit is subject to security clearance, and is mostly granted (about 90 per cent) to married people, while the minimum age for obtaining a work permit is 30' (Berbing 2019, 132).

He does not even mention once that the permit regime described is led by the ISA and that it is the ISA's own policies that have created that situation (Berda 2017a). Another example of the same blind spot can be seen when Berbing describes the desperation of Palestinians in east Jerusalem neighbourhoods which is supported by the 'security vacuum that prevails in the neighbourhoods, where the Israeli security forces and the PA's security mechanisms do not have a deep grip' whilst once again ignoring the fact that it is the ISA that prevents the PA security forces from operating in those neighbourhoods, thus creating the vacuum (Hever 2018). The ability of the ISA to ignore the impact of its own actions on the situation, actions that have led to the outbreak of the violence, is a perfect example of the narrow vision of state counter-terrorism activities that maintains the self-fulfilling prophecy of terrorism and violence (Zulaika 2009).

Who is a Terrorist and What is a Terror Attack?

Israeli law, and Israeli military law were, for many years, reluctant to officially use the word 'terrorists' and used the word 'security offenders' instead. Under this expansive umbrella term, the law gathered a wide range of actions against Israelis by Palestinians, from participating in unlicensed demonstrations to murder, and considered all of them security offences. This wide definition allowed the ISA to arrest, investigate, and charge stone throwers and members of the military wing of Hamas under the same harsh conditions (Ramati and Torn Hibler 2021). In the case study chapters, I presented how the ISA, for the purpose of the AI tool, has defined who is a potential suspect (any young Palestinian man), who is an attacker (any non-Jew that is hurting a Jew or Israeli for being a Jew or Israeli) and what is an individual attack (where the attacker is not connected directly to a terror organisation). As we know, the academic discussion regarding 'what is terrorism' and 'who is a terrorist' is a long and continuous one and big part of the

discussion is dedicated to the question of whether and why it is important to define who a terrorist is (Shanahan 2016, 112; Carr 2007; Englund and Stohl 2016). The answer in the Israeli AI case study, about the need for a clear definition of who an attacker is, is very clear. The AI tool was created to predict the next individual attacker. The first challenge for the developer of such a tool is defining who that individual attacker is, necessary for creating a strong database for training the model (T. B. Munk 2017). One of the most common reasons for AI failures is a non-consistent or non-clear database for the training of the learning model. It is clear then that, in such a complex tool that needs to predict a very specific and unique human behaviour like the will to attack, the training data should be extremely clear. Therefore the definition of who is considered to be an attacker should also be flawless. The data collected for the creation of the ISA tool was based on an analysis of all the possible signals (social media information, communication information, movement and location information, visual information, private historical information and personal connection network information) created by the attackers prior to their attack or attempted attack, during the first three months of the violence (October 2015 - January 2016).[114] According to the ISA data, at that stage, 138 attacks had met the criteria of an individual attack and these involved 149 attackers (some of the individual attacks were actually carried out by more than one attacker). As YS has mentioned, collecting the signals was not enough. They needed to also understand the meaning of the signals in order to train the tool with only the relevant signals and to build the initial model (Y. S. 2021, 68). For this, the ISA attempted to interview the attackers in order to understand which repeating signals were more than just a statistical correlation. One of the challenges was that most of the attackers had been

[114] From interview with retired ISA agent 'BC on 24 April 2022.

killed during their attack or attempted attack (only 22 of them have survived[115]). This meant that a) the ISA had a very small pool of people to interview regarding the attack, considering the fact that not all of them would cooperate with such an investigation and b) the ones who survived were mainly those who had not completed an attack. The need of the ISA to find interviews such as these brought them to interview, for example, NZ[116] - a young Palestinian man who was arrested at an Israeli checkpoint while carrying a knife. According to the facts described in the indictment, he arrived at the checkpoint, took out the knife and put it on the floor and raised his hands, waiting for the soldiers to take him. According to the indictment:

> 'During his police interrogation, the respondent said that he wanted to be in prison, and he would do whatever it took to be in prison. His actions show his determination. For example, in his first statement, after being asked why he wanted to go to prison, he answered "I want to go to prison, period. If you don't allow me to go to prison, I will do something else"'.[117]

The story of NZ is not unique, out of the 22 attackers who survived their attack only five actually harmed someone while the rest were either stopped while advancing on their way to carry out the attack, according to their statements, or they were stopped at a checkpoint carrying a potential weapon. It seems that the psychological profile of the attackers was heavily influenced by interviews with people who did not actually cross the important threshold of attacking. Taking into consideration that the purpose of the AI tool was to exactly identify the person who might

[115] For more about the Israeli policy of using lethal force during that period, please see the Amnesty International report 'Lethal force and accountability for unlawful killings by Israeli forces in Israel and the Occupied Palestinian Territories: https://www.amnesty.org/en/documents/mde15/4812/2016/en/

[116] As NZ was a minor at the time of his indictment, his full name is not disclosed.

[117] The military prosecution Vs NZ, from the author's personal collection.

cross the threshold towards violence, joining together the profiles of actual attackers with those who did not commit any clear act can clearly impact the algorithm by extending its results to those who would never cross the threshold. 'Mixing' such as this, in the definition of who is an attacker, can also be seen in the question regarding what is considered an individual attack. For the purpose of collecting the data for the AI, an individual attack could have been a person pulling a knife in a settlement in the midday light, trying to stab as many civilians as they can, two people shooting a rifle from afar at an army post, or a person driving a car that suddenly accelerates and goes off the road, hitting people on the pavement. All of these different kinds of attacks were part of the initial training bank for the AI model - 49% knife attacks (67 attacks), 39% shooting attacks (54 attacks), and 12% running over attacks (17 attacks).

As discussed in the literature review chapter, many of those attackers and attacks are not defined in the academic world as 'individual attackers' or 'individual attacks'. Just as an example, according to Simon, it is not an individual attack if the attack is carried out 'during popular uprisings, riots, or violent protests' and an individual attack can be defined as such when carried out against the military only 'when the military is not an occupying force or involved in a war, insurgency, or state of hostilities' (2016, 266). Having said that, regardless of whether academic research considers the attacks 'individual attacks' or the attackers 'individual attackers', the research does define very clear differences between the types of attacks and attackers described. Phillips and Pohl differentiate between two kinds of individual attacks according to their relationship to risk - a risk-averse or a risk-taking individual actor (Phillips and Pohl 2012). According to their theory, the profile of an attacker who chooses to engage using a rifle from a distance (risk adverse) is very different from the profile of an attacker who chooses a close knife

attack. Another difference in profiles can be found when it comes to the issue of the level of engagement of the attackers with their surroundings, prior to and during the attack. Pantucci gave a profile and a name to each kind. 'The loner' is a totally isolated attacker moved by self-consumed ideology. The 'lone wolf' is a person who acts by themselves but is supported and engaged by a group of other extremists. The 'lone wolf pack' is a small group of people who are going through the radicalisation process together. And the 'lone attacker' directly operates under the directions of other people (2011, 20–24).

These are just some of the examples of different kinds of profiles and categorisations in the literature that can fit the attacks analysed by the ISA to create the AI tool. It seems that the urgent need for such a tool (according to Sasi Elia, the head of the ISA, he was given two months to build the tool (Shahaf 2020)), led to the ISA putting together in one database, an array of different people who fit different profiles of attack and attackers. This small sample of 149 people included attackers, and non-attackers (people who did not cross the line of violence), risk takers and risk adverse attackers, lone attackers and even pack attackers. The task was, therefore, that the AI tool would learn all the signals collected from the different attacks and attackers, and would know how to identify all of them.

The ISA Tool as Artificial Intelligence

The need for a clear definition in this research of the ISA system is not only required to position it in the academic research of AI but also in the academic research of AI in the field of counter-terrorism or preventive policing. The use of AI in those fields which focus on prevention has, naturally, very clear moral and legal implications and, without such a definition, any discussion about regulation cannot begin to take place (Završnik 2021). As Wang has identified, there is still

no single agreeable definition of what Artificial intelligence is (Wang 2019). Wang suggests that, at this stage, as the research is still very much in its early stages, each researcher should adopt a definition from the ones that exist or create one of their own according to the subject they research (Wang 2019, 29).

On the question of the ISA tool, the issue that arises is a bit different than the regular discussions in the field of AI and that is because of the gap between what the ISA thought its algorithm was doing (and presented as such), and what the algorithm's abilities probably were. Elia Sasi, the head of the ISA cyber unit at the time, described the tool as an extremely sophisticated one that is constantly monitoring street cameras, SMS messages, private phone calls, posts on social networks , phone location, and then send a direct order to the forces to intercept (Shahaf 2020, 4). The tool according to YS, who was the head of AMAN technological unit 8200, has helped to 'prevent tens of lone wolf terror attacks every month' (Y. S. 2021, 76). Considering those statements, it seems like there should be no question that such a predictive tool, which constantly processes a massive amount of data, including visual data, in real time, and which is able to learn to constantly adapt and predict specific human behaviours for a long period of time, and thus can stop terror attacks, is an artificial intelligence tool. For example, Zúñiga, Goyanes, and Durotoye have collected 21 of the latest academic definitions of AI (Gil de Zúñiga et al. 2024, 322–23). When checking each one of those definitions and comparing them to the ISA tool as described by its designers, it fits them all. However, as this research has shown, the actual AI system that was used by the ISA during the period of the research (2016-2017) was probably much less impressive.

It looks like the main power of the tool was to scan the public-facing Facebook posts of Palestinians and rank the users according to their posts, using a few basic search terms that were pre-defined for it. An interesting signal supporting this claim was described by DB and BC in their interviews. According to them, regardless of the situation on the ground, meaning the number of attacks and attempted attacks as well as the tension in the region, the AI tool kept sending the same number of warnings. The fact that the AI tool did not reduce the number of high alerts being generated even when the situation was calmer, questions the algorithm's adaptability, the main feature of a system that has learning capabilities. In addition, as will be presented later in this chapter, it is very difficult to say if the ISA tool was effective even in reducing the number of individual attacks. As such, it is not clear at all if the tool used any machine learning or if it was really predicting anything and therefore might not fit any of the criteria or definitions of an AI system.

Having said all of that, as this research has shown, regardless of whether the ISA tool was a strong machine learning tool with the ability to predict individual human behaviour, or just a basic social media crawler, the impact of the tool was very clear and unquestionable as hundreds of thousands of Palestinians were placed under surveillance, hundreds were approached by the ISA and dozens were sent to jail, based on outcomes of the tool. Therefore, it is my suggestion that, for the purpose of this research and for the purpose of any future research dealing with the potential impact of counter-terrorism or preventive policing tools, deciding whether a tool is AI or not should be based on what its developers thought it should do and not its actual algorithmic abilities. This approach might be helpful when dealing with two different issues: the lack of access and the need for a regulatory framework. As for the lack of access, it is very possible that we will

never get direct access to the ISA tool itself and that this will probably be the case for most counter intelligence and predictive policing AI systems due to the secretive nature of the organisations that operate them. The only information open to researchers will be secondary sources such as reports about the activity of the tool and its impact and interviews with people who worked with it. Therefore, in those cases, those secondary sources are the best we can get and they should guide us. The second reason for adopting a definition of AI is the continuous attempt to regulate the activity of such tools. When an intelligence body declares that it operates an artificial intelligence predictive tool, it is better to define it as such to regulate it, rather than claiming it is not, which allows it to continue to operate unregulated.

Why was the ISA so Supportive of the Technological Solution Provided by a Predictive AI Tool?

Before diving into the question of how effective the ISA AI tool was, there is still the question of the motive to develop and operate such a tool. The investment that is needed to create such a system as described by its designers is considerable and cannot be explained only by a month of violence, intensive as it may have been, and the results of the tool, as will be shown later, were not impressive enough to continue to operate it. The aftermath of 7 October 2023, the day that will be remembered as the worst failure in the Israeli intelligence and counter-terrorism strategy exposed how deep the ISA was invested in technological solutions. The complete failure of the ISA and AMAN in predicting and giving a warning for a complex joint operation of Hamas and the Islamic Jihad which included thousands of people will be studied by the Israeli intelligence community for years, just as the failure to predict the 1973 war is still studied 50 years later. Many details are still unknown and will remain classified for years to come. However, what is

becoming clearer is that the Israeli intelligence community's love affair with technological intelligence tools played a big part in it. In a detailed investigative journalism article, published in *Yedioth Ahronoth*, Israel's most popular newspaper, Ronen Bergman includes evidence from the Israeli intelligence organisations, mainly criticising the former head of the ISA, Nadav Argaman, as the person who advocated for the use of technological intelligence over human sources:

> 'Argaman pushed for the connection between super-advanced technology and operations. Many claim that Argaman had less concern for anything else. Former agency employees claim that the new head simply disparaged the traditional ways of operating the organisation. 'Even in the early days he came to me and told me, "I'm not interested in this biblical espionage,"' says a man who was in a senior position at the time. By 'biblical espionage' Argaman meant the recruitment and operation of agents.
>
> Another former senior official with decades of experience was extremely harsh in his criticism of Argaman and his term of office. According to him, in the case of the ISA, Argaman turned the technology division and the operations division into 'two monsters'. It is not only the size of the divisions, but also the positions that were created in them at the expense of the classic intelligence roles. 'The technology division did not grow by 20-30 percent, but doubled. Each time we saw more positions being moved but where did they come from? From the countermeasures division (that is, mainly from the operators of the human agents and the information collectors alongside the desk officers and analysts responsible for providing warnings and proactive actions against the adversary)'. 'Originally, every part of the agency - the operations, the technology, everything - it was all supposed to help the countermeasures division realise their mission against terrorism,'

> explains the former senior member of the ISA. But the pyramid was turned upside down. Instead of the technology 'helping the countermeasures division' - the technology and operations departments have become the ones that the countermeasures division works for' (Bergman 2024).

This description fits exactly the data from this research about the creation of the ISA preventive AI tool. Argaman, as described by the heads of the technology department, ordered the ISA tool to be developed at an amazing speed when the wave of attacks began in October 2015 in order for it to be up and running by January 2016. From then on, the ISA put its full trust in the outputs of the tool (Barbing and Glick 2019; Shahaf 2020). Argaman himself has repeatedly praised the tool at public events. At one of them, in front of foreign ministers for national security, he said the ISA had foiled hundreds of individual attacks, claiming that:

> 'The iron punch against terrorism is made possible and successful thanks to the combination of quality and dedicated human resources with advanced technology and unique and professional methods of operation. The large investment of the ISA in technological developments in the worlds of big data, learning systems and artificial intelligence, produces a significant leap in the transition from intelligence extraction to forecasting for the purpose of thwarting terrorist intentions and attacks ahead of time' (Saban 2018, 1).

So, what was the source of this fascination with AI technology that led the ISA to make such a huge transformation? Three possible answers come to mind regarding this; one revolves around the concept of Israel as a laboratory for high tech security, the second may be found in theories regarding technological optimism and their relevance to the Israeli Palestinian conflict, and the

third is found in the luring power of the division of moral labour that is offered by such AI technologies.

Israel/Palestine as the 'Lab'

The idea of a 'laboratory', where Israel would use its control of Palestinians to test new technologies that they would later sell, is not a new one. It was presented in the 2013 documentary '*The Lab*' by Yotam Feldman[118] and recently in the book '*The Palestine Laboratory: How Israel Exports the Technology of Occupation around the World*' (Loewenstein 2024). Both the documentary and book present a similar picture of how technology unites the Israeli security forces, the Israeli arms industries and the Israeli political strategy. The more the counter-terrorism technology is used and tested on Palestinians, the more of this technology can be later sold and exported abroad, which of course creates more jobs and more money for the people who have developed it, when they leave their roles in the security services, and more political influence for the politicians who have advanced those relationships. This has been the general feeling among all Israeli leaders in recent years. When former Prime Minister and hi-tech entrepreneur Naftali Bennet said 'when we fight here, we are protecting London, Paris and Madrid' (Harkov 2015), he did not only mean protecting Europe from radical Islam but also how Israel's experiences of fighting terror is teaching others how to deal with terror, an approach that has been well rooted outside of Israel. As US capitol police chief Terrance Gainer said in his visit to Israel: 'Israel is the Harvard of anti-terrorism' (Loewenstein 2024, 14). The creators of the ISA tool were definitely looking for a solution to the individual attacker phenomenon and might have even believed that the answer lay in technology, but they also chose to be interviewed and give

[118] The movie is available online here: https://www.gumfilms.com/projects/lab

speeches about its success. This choice to expose to the Israeli public and the world a secret counter-terrorism tool, leaves no doubt that that the economic and political side of creating such a tool was of importance. A good example for this connection in the specific ISA case can be found in the fact that all three of the founders of the AI tool in the ISA, Argaman, Elia and Barbing, are now employed by leading cyber security companies offering similar products.[119]

Technological Optimism and the Israeli/Palestinian Conflict

Another possible answer to the question of why the ISA became so attached to technological solutions like the AI tool can be found in the theories of technological optimism. Technological optimism is a doctrine that developed in the 1970s/80s mainly regarding environmental issues such as food production and energy sufficiency. It claimed that putting effort into advancing technology would solve all of the problems arising from the growing needs of a growing population, an approach that was, and still is, highly criticised in academia both as being without basis and as a tool that ignores the difficult decisions to be made regarding the use of energy and population growth (Basiago 1994; Gonella et al. 2019). A similar kind of optimism can be found globally regarding the use of AI in counter-terrorism, and it also attracts similar criticism. The optimistic view that technology, and especially AI technology combined with advanced surveillance technologies, will provide protection against terror attacks is, unsurprisingly, very popular in technological security companies and security agencies and highly criticised in academia as being another way to focus on and only deal with the symptoms, whilst ignoring the root causes of the violence (Mahon 2022; Avis et al. 2025). This then should be no different when it comes to the ISA's predictive tool. Publications by, and interviews with, the tool's creators and

[119] Barbing and Elia are employed by 'Blackwall Global' and Argaman is employed by 'Cognyte Software'.

developers in the first couple of years after the tool was activated, are all filled with optimistic views, that the issue of individual attackers in Israel would be solved and, if it is solved using technology, everything can be solved using technology (Berbing 2019; Shahaf 2020; Y. S. 2021). In my opinion, this optimism, in the context of the Israeli/Palestinian conflict, is not only based on Israel's belief in its technological powers but also, very much like environmental technological optimism, on the pessimism around finding another solution to the constant violence between Israelis and Palestinians. Just as environmental technological optimism is very much driven by the realistic but pessimistic view that humanity will not manage to reduce its dependency on fossil fuels and reduce its environmental footprint, the optimism of the Israeli security forces was driven by a growing pessimism around whether a political solution for the Israeli/Palestinian conflict is at hand, a pessimism the has become more widespread in recent years (Evans 2023).

The Power of the Division of Moral Labour

As discussed in Chapter Three, technology and artificial intelligence specifically can provide a strong incentive for those who are dealing with difficult moral decisions to delegate their moral dilemmas to a machine (Salatino et al. 2025; Feier et al. 2021). The activation of the ISA AI tool is a great example of such delegation. The AI dramatically reduced the painstaking job of ISA analysts trying to identify a needle in a haystack, and it brought them clear candidates with a risk score attached. As BC describes, their main job was: 'to decide which of the people the tool was ranking were ranked high by mistake - the rest we would just approve for one of the actions'.[120] This change, from looking at a vast amount of data and trying to identify people who potentially could be attackers and acting against them, to looking at a list of potential attackers and deciding

[120] From an interview with retired ISA agent 'BC' on 28 April 2022.

which one of them might not be an attacker, had another impact on the agents and this was greater than merely reducing the amount of manual work. Prior to the launch of the AI tool, the ISA agents were burdened with a moral duty to decide, on their own, which young Palestinian, who had not yet done anything, was a potential attacker that needed to be approached and perhaps sent to prison. The AI recommendation system flipped the picture, as the AI tool took away the moral burden of predicting who was going to be an attacker. This time it was 'science' or 'hi tech' that gave the answers, leaving the ISA agents to choose who wasn't a potential attacker from a list of potential attackers. In a way, instead of condemning people who had not done anything wrong yet, they were now saving a few of them from that fate. Most of the moral burden was now in the hands of the machine. Another example of this kind of division of moral labour between an AI recommendation system and Israeli military operatives could be seen during the first days of the war between Israel and Hamas after the 7 October attack. According to a report by the Israeli NGO, 972+, Israel used an AI tool to automatically generate thousands of targets to be attacked in Gaza based on digital signals collected by it. According to an Israeli official:

> 'From the moment this machine was activated, it generated 100 new targets every day. You see, in the past there were times in Gaza when we would create 50 targets per year. And here the machine produced 100 targets in one day'. Another user of the AI added: 'I would invest 20 seconds in each target at this stage and do dozens of them every day. I had zero added-value as a human, apart from being a stamp of approval. It saved a lot of time' (Avraham 2023, 4).

However, the target-choosing AI tool had another impact, apart from dramatically raising the number of targets and approved attacks. In a follow-up article on the 972+ report, published in The Guardian, an Israeli official was quoted as saying that they had more faith in a 'statistical mechanism' than a grieving soldier. 'Everyone there, including me, lost people on October 7th. The machine did it coldly. And that made it easier' (McKernan and Davies 2024). Those two examples of AI use by the Israeli security forces, where the proactive moral choice is given to an algorithm leaving only a passive role to the agent, fits into the driving power of the theory of division of moral labour. As Scheffer identified moral responsibility is perceived as relative, for example:

> 'We have greater responsibility for what we ourselves do than for what we merely fail to prevent… One's primary responsibility, in general is to avoid harming others oneself' (Scheffler and Munoz-Dardé 2005, 231).

Giving the AI the decision of causing harm, arresting a person, or even killing them, and leaving the agent with only the preventive role of approving or not approving, is a huge lifting of moral burden held by the agents before the algorithms were activated. This also corresponds with the finding on the lack of responsibility agents feel about decisions made following suggestions by algorithms (Salatino et al. 2025) and how others perceive it as reduced responsibility (Feier et al. 2021). All these together create a strong incentive for agencies to develop such tools, for agents to use them, and even more so for agents not to question the tool's results and effectiveness.

Among the three different reasons explaining why the ISA was so enthusiastic in adopting the AI system, the division of moral labour is, in my opinion, the most dangerous. The decisions made by counter-terrorism agencies often have difficult moral implications. The moral dilemmas

involved in those decisions are important and should be a part of the decision-making process. As we are very far away from, and maybe will never get to, a place where artificial intelligence has an independent moral compass, those dilemmas should stay in the hands of humans, no matter how uncomfortable it is. Whilst it is understandable that law enforcement and counter-terrorism agents wish for a statistical tool that will help them skip over difficult moral decisions, at the moment, leaving those decisions in the hands of AI is the same as leaving those decisions in the hands of a diagnosed psychopath.

It seems that that none of the reasons why the ISA continued to praise their predictive AI tool, even though its effectiveness was in doubt, are good reasons for supporting the implementation and development of future tools like this. It is actually quite the opposite. Those reasons, gaining financial and political power, being immersed in technological optimism in order to ignore political solutions, and diverging from difficult moral decisions, merely strengthen the arguments against the development and use of these tools. Therefore, the fact that counter-terrorism organisations around the world are using or actively looking to develop such tools, does not in any way reduce the identified costs and dangers that arise from using them.

Could the ISA Tool Predict Terror Attacks?

Before diving into specific data signals and evaluating the specific ability of the ISA AI tool to identify an attacker before an attack, it is important to briefly discuss the theoretical ability of the ISA AI tool to predict an individual attacker. As discussed in Chapter Two, the current literature points to many possible challenges for person-based predicting, and especially person-based predicting in counter-terrorism. For a quick recap, most of the literature regarding person-based policing has identified the issues of lack of quality initial data, bias, and ability to validate

the findings, as the main problems of algorithms used by police forces to pre-identify criminals (Ferguson 2017; Babuta 2020; Hung and Yen 2021). Adding to those challenges are the specific challenges identified in the literature regarding creating an algorithm that could pre-identify a terrorist. First, as Monahan has identified, terror attackers and attacks are so varied it is extremely difficult to create a single psychological statistical model that will fit them all (Monahan 2012). Second, the fact that terror attacks are a rare phenomenon makes it almost impossible to collect enough data for a predictive algorithm (Munk 2017). And third, the amount of surveillance that is needed in order to collect enough signals make it very difficult to do it legally (Schröter 2020; Fernandez and Alani 2021).

Having said that, the lack of access to data from counter-terrorism organisations that use these preventive measures means that we cannot come to the unequivocal conclusion that such tools cannot be effective or that the ISA tool is not effective at all. In its support, the heads of the Israeli security forces swear by the tool's effectiveness (Y. S. 2021; Shahaf 2020; Briner 2018; Berbing 2019), all claiming the tool was a great success. Another aspect that supports the claim that this specific ISA AI system was successful in predicting attackers is the specific conditions within which the ISA system operates.

First, regarding the problem of creating a psychological statistical model of a terrorist, the relatively narrow scope of attacks and attackers - young Palestinians from a similar geographical area following a similar ideology - might solve some of the issues as identified by Monahan (2012). Second, regarding the statistical problem as raised by Munk (2017), the specific case study of the ISA AI presents probably the best statistical scenario for counter-terrorism AI. Unlike other big intelligence bodies who are looking to find an attacker out of millions of possible targets, the

ISA has limited the pool of potential attackers to only 200,000 men in a specific age range. The data pool of attackers that was used to train the AI model was also relatively large especially in comparison to other countries where such attacks are not as common. During the three months prior to the activation of the tool, the ISA collected more than 100 profiles of individual attackers. This data together presents a ratio, between the pool of potential attackers and the number of cases for the AI to train on, that is much better than the ratio used by Munk (2017) to reach his conclusion of the ineffectiveness of such a predictive tool. Third, regarding the issue of bias in data and results, this issue is not relevant to the ISA AI tool, as the tool is designed from the start as an inherently biased tool. It was looking for Palestinian young men from a specific region, and the data that was used to train it was based on cases from the same population, so the challenges of unconscious bias are not relevant to it. Fourth, as for the questions of legality, the surveillance was carried out by the ISA on Palestinian residents of the OPT, and as mentioned in Chapter Six, was carried out under the applicable law according to Israeli Supreme Court rulings. This surveillance collected a wide range of signals (e.g. online activities, phone locations, communication signals) which can rarely be collected legally in bulk in democratic countries.[121] The ability to collect such a wide variety of signals in bulk and then use some of them as evidence in court without being challenged answers some of the issues raised regarding the possibility of using these tools from the legal point of view.

Summing-up, the current research does not support the claim that it is possible for an AI tool to effectively predict the identity of a potential individual attacker. Having said that, looking at the

[121] It is important to make clear that the term 'legally' refers here to 'legal under Israeli law', and not by any standards of international law or human rights law.

specific conditions the ISA tool operated in, it cannot be refuted that it is at least theoretically possible that this kind of tool could have been effective.

Was the ISA Tool Effective?

The question about the effectiveness of the ISA AI tool sounds simple on the surface of things: did it work or not? However, if we dive into it, it is obviously not that simple as it includes many sub questions such as 'what is effectiveness?' 'Does it stop a person from committing an attack?' Does it reduce the specific wave of violence? Does it reduce the level of violence in general? How do you measure it? What data do you have? For what period? In comparison to what? The problem of measuring the impact of a preventive counter-terrorism tool or strategy is a known ontological and epistemological one. When it comes to the short-term impact, in our case, did the tool identify a person that was going to attack and did the ISA intervention stop it? Evaluating effectiveness means adding in the regular issue of proving the causation of events that take place due to specific circumstances. The even bigger challenge is proving the causation of events that don't take place due to specific circumstances (Munk 2017). As for the long-term impact, in our case, proving this tool has helped in reducing the phenomenon of individual attacks means knowing how to differentiate the specific actions of the AI tool from all other aspects that could have resulted in such a decrease (Jore 2021). Understanding if the tool has made a general impact on violence in the region, makes it even more difficult to measure. Measuring those in a perfect data world would be very difficult. Measuring those in the very limited data world of security and counter-terrorism is almost impossible (van Um and Pisoiu 2015).

As for the question of measuring the effectiveness of this ISA tool, the answer is the same. There is no way to provide a clear answer as to how well the AI tool worked. Regarding the short-term

impact, as the AI tool is supposed to identify an attacker prior to the attack and activate some chain of events that will prevent such an attack from happening, we cannot know if the supposed attacker would have carried out the attack or not. The head of the ISA at the time, Nadav Argaman, claimed that the tool helped in the prevention of hundreds of attacks (Briner 2018). However, it is not clear how the ISA calculated this. They may have counted every interaction the machine initiated as a prevented attack. As the ISA is not a research body, it is unlikely that they used a control group of cases to check if the AI tool identification was accurate. As for the long-term impact, we can look at the number of individual attacks following the activation of the tool but, as stated, it would be very difficult to separate any change from all the other political, social, and other security efforts happening at the time, and the same can be argued about the violence in the region in general regardless of the kind of attacks. Having said that, there are some very interesting signals regarding the short-term impact and the long-term impact of the AI tool that cannot be ignored.

Measuring the Short-Term Impact of the ISA AI Tool - Identifying an Attacker Prior to an Attack

The most clear signal regarding the effectiveness of the AI tool in identifying an attacker prior to an attack is that none of the 104 Palestinians who were arrested, questioned and indicted, having been identified as a result of the ISA tool, had a single collaborative signal in their file to support the theory that they wanted or planned to carry out an attack. As described in Chapter Six there was no letter of intent, no collaborative witness by a friend or family member on such intent, nor weapon found on any of the people who were identified and arrested following identification by the AI tool. This finding is quite striking considering that BC and DB described how, out of the

people who were identified by the tool, only those whom the ISA felt there was no other way to stop (via a warning or a phone call to the family) were arrested, meaning the ones who were the closest to carrying out an attack. It is even more striking when taking into consideration the high proportion of confessions usually obtained by the Israeli police and the ISA agents in the OPT – a 96.8% rate of confession in ISA investigations and 62.8% in police investigations (Ramati and Torn Hibler 2021, 479). This research about police and ISA confessions has also found strong indications of a clear presence of false confessions being obtained by the ISA because of its harsh methods of investigation. It is startling therefore that none of the people arrested following the recommendation of the tool have confessed to having planned to attack.

I do not have a very good explanation for this finding, except the obvious one: that the ISA AI recommendation system did very poorly in identifying potential attackers. Having said that, there are some other possible explanations. Firstly, such a confession - planning to attack people - is not a very easy one to obtain (although, as mentioned above, the ISA is very good at obtaining confessions for much more severe and complex offences). Secondly, the investigators felt they had enough incriminating evidence in the shape of the posts written by the people arrested so they did not push for such a confession. It is true the some of the posts carried worrying message such as 'I am the next martyr' which might be considered by the ISA as a confession to plan an attack (although, as shown in Chapter Six, the court did not find this evidence compelling enough in some cases, so an official confession would have been necessary). Thirdly, the ones who did confess to such a plan were subject to administrative arrest, in which cases I do not have access to the case materials (although the court found administrative arrest to be a possibility only when a criminal case wasn't possible and a confession would make it a very easy criminal law case).

Those explanations are possible but, as described, not extremely credible. They might explain a low number of confessions about the intention to attack but it would be very difficult to explain not a single confession.

Another possible explanation for the fact that none of the people arrested - on foot of a recommendation by the AI tool - claimed they were planning an attack, is that the AI tool was so powerful that it knew how to identify these potential attackers even before the thought had crystallised in their mind. An indication of this can be found in the interview with Elia Sason (the head of the ISA cyber unit) where he explained the power of the AI tool: 'from 70 likes on Facebook, I know how to appreciate what you have in your subconscious, I know more about you than you know about yourself. Do you know how powerful that is?' (Shahaf 2020, 7). This might suggest that the ISA were very confident in their tool predicting who was going to be the next attacker, to the level of predicting it before such a decision or plan was conceived consciously. This kind of explanation is of course very difficult to refute due to its tautological nature - all of the potential attackers arrested did not say they were going to attack because they did not yet know they were planning to attack.

In conclusion, regarding this issue, there are two possible explanations for the fact that none of the people arrested following the AI tool recommendation had any collaborative signal to support the fact they were about to commit an attack. One is that the AI recommendation system failed in identifying potential attackers. The second is that the AI system was built to identify the attackers prior to them planning or thinking of such an attack. As discussed previously, regarding the current research position on the possibility of any AI recommendation system identifying attackers, I lean more towards the first explanation.

Measuring the Long-Term Impact of the ISA AI Tool - Reducing the Number of Individual Attacks

Another way to try and understand the effect of the AI tool is to look at its impact on the phenomenon at large. Did the activation of the AI tool lead to a reduction in the number of attacks? The first issue you encounter when asking that question is the famous question of causation. As described in the case study chapters, the security forces have tried several tactics to reduce the number of attacks. Improving the accuracy of their intelligence was just one of them. For example, finding and arresting the main inciters online, strengthening cooperation with the Palestinian authority, and punishing the families of specific attackers by withholding work permits (Numa and Liraz 2019, 134) - any of these tactics might have been the one to have brought the wave of violence to an end and some of the military commanders during that period are sure it was actually these other tactics that did it (Goffman 2019; Carmeli 2019). However, if accepting the ISA's assumption that the AI tool was able to identify potential attackers before they even planned their attack, then its activation should have had a clear impact on the number of attacks and attempted attacks carried out. However, when we look at the number attacks and the timing of when the AI tool was activated, we don't get a very clear answer.

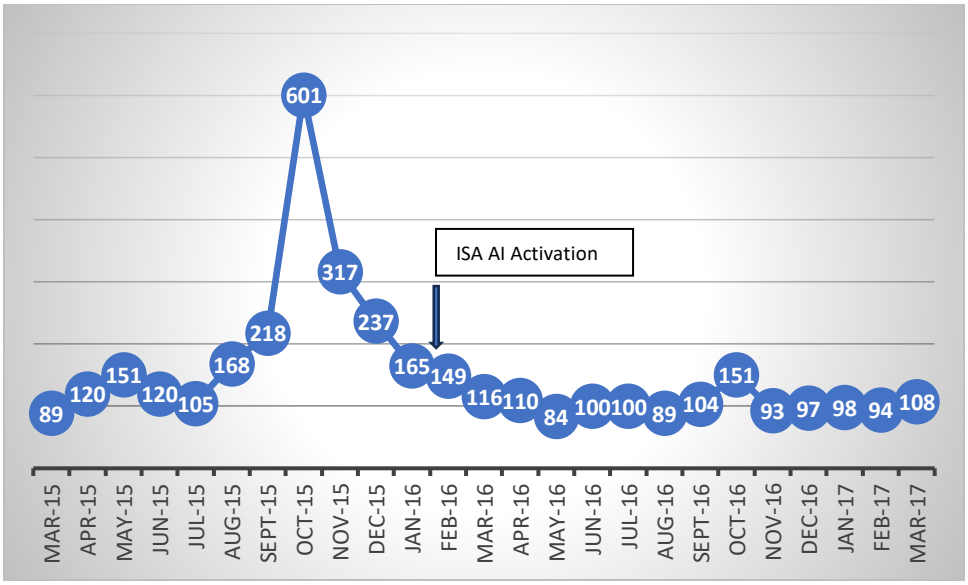


Figure 12. Attacks by month March 2015 to March 2017

As shown in *Figure 12*, the number of attacks, which reached its peak in October 2015, started to decline before the AI tool was activated around mid-January. A simple unpaired t-test comparing the periods before and after the activation of the AI tool also provides no statistical assurance that the activation of the AI tool had any impact on the number of attacks carried out.[122] Moreover, when looking at the waves of individual-attack terrorism that followed in IL/PS in July 2017 and October 2022, they look very similar in their statistical behaviour to the October 2015 wave.[123] A spike of concentrated individual attacks ignited due to an Israeli attempt to change the status quo in Al-Aqsa followed by a declining in the attacks after a wave of arrests. The fact that, after the AI tool had been activated, similar spikes were taking place without the AI managing to predict the attacks and attackers does also not support the claim of a high degree

[122] The two-tailed P value equals 0.1616. By conventional criteria, this difference is considered to be not statistically significant.

[123] See the statistics for those months in the ISA monthly reports of attacks at https://www.shabak.gov.il/reports/.

of tool effectiveness. Also when examining the general levels of violence between Israelis and Palestinians, there is no evidence to support that the activation of the tool and its impact on the activity of Palestinians on social media, reduced the levels of violence, as the average number of all Palestinian attacks, as collected by the ISA, remained similar in the following years with spikes in 2018, 2021, and, of course, 2023.[124]

Having said that, in the same way that finding a statistical correlation does not immediately point to causation, the reverse argument is also possible. The fact that there is no clear statistical correlation does not mean causation does not exist (Anscombe 1973). This argument is even stronger when discussing the effectiveness of counter-terrorism tools that usually deal with relatively small numbers - where the power of statistics is less clear. But more than that, given that a t-test, a tool for scientists to set a standard for correlation, sets the number for scientific proof validation at $P < 0.05$ and given the statistical result for the correlation of the AI tool activation to the number of attacks was $p = 0.16$, it might mean something very different to counter-terrorism agencies than to a social scientist. In the eyes of a counter-terrorism agency, the fact that there is a 16% chance of there being no correlation between the tool and its impact means that there is an 84% chance that there *is* a correlation between them and that is very promising regardless of the 0.05% line drawn by social scientists. Since 9/11, many counter-terrorism agencies have taken this approach - that there should be almost no limit to 'the costs of saving one innocent life' (Barnes 2012, 1633).

Therefore, the question of AI tool effectiveness in the eyes of the ISA might be completely different to the analysis of its effectiveness up to now, as DB described well in his interview: 'In

[124] Yearly statistical reports by the ISA, as found at https://www.shabak.gov.il/reports/

my opinion, if, out of those thousands of people we approached there was one who could have committed an attack and did not, then it is great, or even just because people now know that they cannot post whatever they want online, and therefore one person has not been brain washed to attack and an attack did not happen - that is also great'.[125]

To conclude the question of effectiveness, it seems that, although that specific conditions that were available to the ISA in creating and operating the tool were optimal for creating such a tool, and although the ISA leaders praise the effectiveness of the AI tool in helping to prevent hundreds of attacks through identifying attackers before they decide to attack, the data available does not clearly support this. However, it seems that even the more sceptical ISA agents, such as the retired agent DB, see some benefit in the AI tool in stopping the wave of individual attacks, if not through identifying possible individual attackers, then more as a general deterrent against Palestinians using online posts to incite and this, according to him, helped to generally reduce the level of online triggers available to Palestinians.

Taking into consideration that the AI tool was created in 2016, the AI capabilities that were available at the time, the specific need for the AI (identifying a specific person from a group of people before they decide to attack) and all of the signals that have been collected as part of this research regarding the abilities and effectiveness of that AI, it is safe to assume that the recommendations of the ISA AI tool were not much more effective than any other random sample that could have been taken from a targeted population that fitted the initial profile. To be clear, I believe that if the ISA had, during that period, randomly approached 2,000 young Palestinian men from the OPT who were politically active on Facebook and had treated them in

[125] From interview with retired ISA agent 'DB' on 26 April 2022.

the same way as those who were identified by the AI tool (threatening calls to them and their families and some arrests), the impact on the level of violence and the number of attacks would have been very much the same, if there was any impact at all.

Conclusion

In this chapter, I have tried to answer two questions: What can we learn from the reasons for the design activation and operation of the ISA's AI tool? Was this tool effective in any way, short or long term?

I have tried to expose all the intentional and unintentional blind spots and gaps that follow on from the initial assumptions made by the ISA and the development of the tool. I have discussed the reasonings that might hide behind the ISA's fascination with technology, I have presented the gaps between the Israeli security forces' bombastic declarations regarding tool effectiveness, and the findings this research has collected regarding the tool's abilities and impact.

In the next chapter, I discuss the conditions under which it may be possible to develop a more effective predictive tool, as well as the legal circumstances under which such a tool could operate within liberal democracies. I then address the normative question of whether such a future tool—operating under the assumptions outlined above—should be used at all.

Chapter 8 – Can and Should We Use AI to Predict Individual Attackers?

In this chapter, I move beyond the specific ISA AI tool to explore the theoretical possibilities of using preventive AI tools to identify individual attackers in counter-terrorism. I address two key issues. First, building on the analysis from the previous chapter, I outline the optimal conditions under which a theoretical predictive AI tool could assist in more accurately identifying potential attackers from a statistical standpoint, while remaining legally compliant within liberal democratic regimes. In the second part, I examine the potential impact of deploying such a tool through the lens of Lindahl's model of counter-terrorism (Lindahl, 2019). The tool is assessed using the model's five components—key assumptions, basic principles, strategies and tactics, prevention, and evaluation—to determine whether, under optimal conditions, it aligns with the principles of critical counter-terrorism thinking.

What Would It Take to Have a More Efficient AI Counter-Terrorism Predictive Tool?

The ability to predict a terrorist attack before it happens has been has been a major aim of counter-terrorism and police forces around the world and many attempts to build these algorithms have been and are being made around the world (McDaniel and Pease 2021, 2). It seems that the prize at the end of the race of creating such a functioning system is so high in the eyes of those organisations that not a lot of thought is given to the impact and meaning of building and activating such a tool. The continuous race to develop such a tool, as described above, of course exists because many agencies see how AI is incorporated into everyday life and their cyber units feel obligated to try and harness it for their purposes (Babuta et al. 2018). AI

abilities are advancing at an amazing speed right in front of our eyes, even during the period of this research. AI tools have developed from tools that have been quietly working in the background to become something that provides everyday assistance to online users. Therefore, it is not improbable at all to assume that, in the very near future, a tool like the ISA tool will be able to offer warning lights about potential terror attackers and attacks at least a little bit better than a coin toss. In the following pages, I will analyse what it takes to create such a tool.

The Training Data for the Tool and the Issue of Bias

To create a working machine that predicts human behaviour, it must be provided with clean and clear training data where it can try to identify patterns that are invisible to the human eye and brain. This initial training data is created by humans and therefore must deal with one of the most difficult challenges in creating a dataset for AI, the problem of bias.

We as humans are, as many social experiences have proven, drawn to bias like a moth to a flame. Our natural tendency is to believe that what we just now saw happening will happen again in the same way, even if it is statistically meaningless. We are drawn to clear correlations without searching for clear causations (Blair et al. 2020). If we look at cars that break the speed limit and, out of ten of them, three are red cars and the rest are all kinds of colours, our brain will immediately direct us to follow the red cars more closely, creating a whole meaningless story about how the drivers of red cars are less careful than other cars. The same can happen with data we choose to enter into an AI tool. Choosing the 'relevant' data to enter into such a machine is filled with conscious and unconscious bias (Mayson 2018).

A clear example of this bias can be found in the training data that was entered into the ISA tool. The first clear and consciously biased decision was to only focus the tool on following Palestinians

and only those without Israeli citizenship in the OPT and East Jerusalem. Therefore the data training set was only based on these cases. This is not a bias created by the blindness of the ISA, as the ISA is using its Jewish division to deal with violent and deadly attacks by Jewish settlers on Palestinians (Eiran and Krause 2018). The attacks, which take place at the same time and in the same geographic space as the Palestinian ones, are not organised by a distinct terror organisation and can be perceived as individual attacks according to the same criteria created by the ISA for the individual Palestinians attacks. In terms of their immediate impact, the heads of the ISA have identified them as extremely dangerous, not only to Palestinians, which they are less worried about, but also to the whole Israeli Jewish population as it quickly inflames a counter reaction (Borger 2024). Having said that, the ISA has clearly and consciously decided to exclude those attacks and attackers from the goals of the AI system and therefore from its training set. The reasons for such a choice could be varied. The ISA does not perceive Israeli attackers as terrorists. The ISA thinks that it has different legal responsibilities when dealing with the rights of Israeli citizens in comparison to the lacking-in-rights Palestinians. The ISA does not have enough data about Israeli attackers due to their very low success rate in solving settler violence cases (Sharon 2022). Regardless of the reasons, the impact of such an obvious bias is clear: the ISA AI tool is another Israeli mechanism which forms part of the apartheid regime in the OPT and which differentiates between Palestinian and Israelis under the same circumstances, a regime that is, even as the ISA has identified (Berbing 2019), one of the reasons for the rise in Palestinian violence. The same dilemmas can arise in any country that wants to operate such a predictive tool. Violent attacks have never belonged to one political, religious or ethnic group, and as nation states become less homogenous, it is becoming even clearer. A threat by an individual who

decides to attack others can come from very different directions and therefore all types of attackers and attacks should be incorporated in the database.

Other kinds of bias are less visible and are based on the investigators' misconceptions that 'may replicate (and in some cases amplify) the existing biases inherent in the dataset' (Babuta 2019, 12). A clear example of this can been seen in the nature of the training data that was entered into the ISA tool. As we saw earlier in this chapter, the ISA based a substantial amount of the training data on attackers who had not committed attacks but who had been arrested at friction points in possession of a potential attack weapon, usually a knife, some of them stating clearly that their only intention was to get arrested. Including all those cases in the training data could definitely have created a bias that skewed the AI results towards looking for the wrong kind of subjects. This example shows why the problem of bias in the training data, which exists in any AI tool dealing with human behaviour, is much more complicated in a predictive tool that is supposed to identify terrorist attacks among the cases. The low number of attacks that are defined as 'terrorism' makes the impact of any small bias in the choice of training cases statistically dramatic. Entering two wrong cases into this kind of dataset might be enough to throw the AI totally off course.

Another aspect relating to bias and the training data is the issue of bias in sources. Unexpected human behaviour is very complex and it can be affected by a variety of factors (Lalljee et al. 1982). Choosing to enter some factors into the training data and exclude others also creates a bias in the system as it will then lack information that might be important in creating a prediction. In the ISA tool for example, the ISA has chosen to collect what it sees as the most important data: family status, online behaviour on all accessible platforms, and patterns of movement based on phone

locations and face recognition cameras. Whilst it is an impressive amount of data, so many other aspects of impact human behaviour are lacking, for example, nutrition has been proven to have an immense impact on decision making (Strang et al. 2017). So too does the weather (Keller et al. 2005) or even school grades (Lettau 2021). When it comes to deciding whether or not to attack someone, the impact on a young man of factors such as having a slight fever, not eating all day, receiving a bad grade in school and walking home in the pouring rain, could be much more dramatic than if he liked or did not like a social media post praising a previous attacker.

Summing-up on the question of the training data, in order for a predictive counter-terrorism AI system to develop a relatively effective algorithm, the dataset for its training should include as many as possible relevant cases and only clear and obvious ones and, for each relevant case, the data to be collected should include every possible data source available without imposing on the AI only the data that the developers think is relevant.

Tool Monitoring and the Issue of Rights

The issue of the data needed for the AI training model brings us directly to the issue of the everyday data needed for the model to make its predictions. As discussed above, the amount of private and personal data needed to create a somewhat successful predictive tool for a complex and unique human behaviour such as an individual attack, is endless. Whilst we as a society might accept that, in order to understand the motives of violent attackers, counter-terrorism agencies have to collect data regarding attackers daily lives prior to their attacks, we are much less willing to accept such data collected twenty-four hours a day on an innocent population, and even much less on ourselves. In recent years, both in the US and Europe, most of the attacks that are defined as terrorist attacks are committed by their country's own citizens (Frostenson 2015). Operating

these tools successfully in those countries means collecting an enormous amount of their own citizens' private data, which leads to a very serious legal and constitutional problem. As Wall stated:

> 'ML tools hold much promise for countering domestic terrorism, but caution is warranted. In spite of the vast experience the United States has in fighting terrorism abroad, the tools developed for battlespaces like Afghanistan or Yemen are not necessarily optimal in a domestic context…. At home, the U.S. government is not at war with anyone, even if extremists routinely declare war against the state. This means that any machine-learning powered counter-terrorism policy must align to a constitutional framework, something that will curtail the most ambitious efforts to use ML for threat detection within social media' (Wall 2024, 2).

The same constitutional rights, which protect privacy and free speech can be found in some form or another in many other countries around the world, and if those rights limit the surveillance of social media postings, they will of course forbid the constant surveillance of other signals such as the monitoring of phone locations, facial recognition cameras, credit card transactions and medical information, all of which are needed to create a successful tool to predict human behaviour. This hurdle is not an easy one to overcome, and it means that such a tool could be activated in only one of three possible scenarios. First, where the power that activates it has no duty to maintain the right of the population to privacy or free speech such as with the military occupation of a different nation or in a dictatorship. Second, when the power that activates it has legal duties to maintain those rights but activates the power illegally without telling the citizens of the country (an unfortunately common situation in many counter-terrorism

organisations around the world). And third, when the people of that nation willingly revoke those rights in order to identify violent attackers and prevent attacks.

Regarding the first scenario, in appalling situations like a military occupation or a full-blown dictatorship, the discussion on the right to privacy and freedom of speech should always be linked to other, more important rights that are usually also infringed in those situations, like the right to life, the right to freedom of movement, the right to property etc. The importance of the right to privacy and freedom of speech in dire situations such as those is different as it is clear, for example, that the Palestinians living in tents on the streets of bombarded Gaza are less concerned right now about their right to privacy.

Regarding the second scenario, although it is an unfortunately common one, it is not a sustainable one as, at the end of the day, it will always be discovered. A clear example is the US NSA surveillance program which was ruled illegal by the US courts after Edward Snowden exposed it (Reuters 2020). The involvement of humans in the process, either as part of the team that design and operate the tool or the people affected by it, will always lead to its exposure in the end.

It seems then, that for most democracies, the only possible option for activating such a tool successfully, is to present its potential benefits to its population and create the legal mechanism that will allow its activation. This scenario might sound unreasonable but, as people get more and more used to the presence of AI technologies in their daily life, I do not think it is impossible that, with the right assurances such as that the data collected is only accessible to the AI system and cannot be accessed by humans, people in some countries will be willing to accept this form

of surveillance in order to fight terrible phenomenon like, for example, the rise in school shootings.

The Implications of Being Identified as a Risk by Such an Algorithm

It is obvious, and even expected, that no predictive AI tool of human behaviour will ever be extremely accurate regardless of the amount of data gathered by it. No AI will know how to calculate the impact on the decision to attack of a facial expression such as the frown of a parent or a partner, or the impact of a surprise hug by a friend or even just a smile from a stranger on the street. The most optimistic possibility is that such a tool could identify, with some statistical accuracy, the people who have highest potential to become attackers. Some might attack, some might never.

Reaching the stage where a functioning and legal AI predictive tool is somewhat accurate in predicting who might be a potential attacker, is only one issue when it comes to operating such a tool. The second relates to what kind of actions could be taken after such an AI tool statistically identifies a potential attacker.

The problem here not only lies in the lack of criminal evidence against those people but it is also connected to another inherent problem with any existing AI recommendation system, the 'black box' problem. The idea of the black box is that, because AI can deal with and process a massive amount of data, it can find answers in places we do not understand and so the decisions it makes are sometimes not clear to humans, and the AI itself cannot explain them. A simple example of this can be found in Alpha Go's famous 'move 37' in its second game of Go against the reigning world champion of Go. It was a move that no human could understand and predict and was considered a mistake until it was proven that it led to victory. The problem was that, at the time

of the decision to move, there was no one who could explain it, especially not Alpha Go which was not programmed to explain its decisions (Metz 2016). Whilst not explaining a Go move is maybe frustrating but not dangerous, it is obvious that making decisions regarding human life need some kind of explanation. In such a scenario, an explanation is needed not only to learn from the AI decision but also to monitor its accuracy. In a famous AI test, an AI system was given a large dataset of images with descriptions and then tested by being presented with a new set of images to identify. A re-occurring strange result was the AI describing certain pictures that included a specific kind of fish, although fish had not been included in the images. Only by reverse engineering the machine's decisions, was it discovered that the set of training data for that fish often included the fingers of the fisherman that was holding it as a trophy. The AI identified the fingers as a good signal for this image and therefore when it identified fingers in an image, it led it to a definition of a fish (Brendel and Bethge 2019). This kind of mistake can easily happen in any set of training data, and without the ability to explain the decision, there is no way to fix it. The lack of such an explanation mechanism is one point that is highly criticised. An example is in the Liberty report about predictive policing:

> 'If no one can explain how a decision about you has been made, it is impossible to challenge it properly. This is a significant problem given that, where the state infringes on a person's rights, there needs to be a legal basis for that infringement. This is to protect people from being treated differently based on arbitrary or illogical factors such as where they live' (Couchman 2019, 42).

The ability to present a statistical explanation is important, not only for the accountability of the decision, but also as another check on the risks of using biased data. Returning to the example of

the AI identifying a fish, if the AI could explain that it was using fingers as an identifier, it would have been much easier to fix. This issue of accountability in predictive AI tools has become the focus of a growing area in AI research that seeks to develop explainable AI tools that 1) produce more explainable models while maintaining a high level of learning performance (e.g. prediction accuracy), and 2) enable humans to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners (Barredo Arrieta et al. 2020, 3). It seems, with the growing pace of AI advancement, and its language and visual abilities, predictive AI tools could present some kind of understandable statistical explanation of their recommendations which will help with the black box issues. However, a statistical explanation, as good as it might be, would still be just a statistical explanation of how a person might in the future take part in a violent act.

With the ISA's predictive tool, the recommendations were sent to a human agent, who would assess them according to the information already collected on the individual and decide to merely keep the individual under a higher level of surveillance, to make a phone call to the individual or his family and warn them that the ISA is aware of their activity, or arrest them, all according to the level of risk identified.[126] The final decision then was back in the hands of humans, with all the regular bias they bring with them.

It is obvious that, even when we have a somewhat effective predictive system, most democratic countries will be extremely hesitant to arrest their own citizens based on such a prediction, and this kind of measure will only be used when clear data, analysed by humans, shows an imminent

[126] From interview with retired ISA agent 'DB' on 26 April 2022.

attack is on its way (it is important to emphasise again that in non-democratic situations such as wars, or non-democratic countries, people will probably be less hesitant (e.g. in the Israeli case)). The more likely solution, when there is a recommendation from an AI tool, will be some kind of even more enhanced surveillance on the individuals identified, together with the incorporation of existing 'de-radicalisation strategies'. This catch-all term, as identified by Lindekilde, brings together 'policies that differ widely in terms of aims, targets and involved actors' (Lindekilde 2016, 255). Most of the strategies have been highly criticised, however, there is no doubt that they are better than a more violent solution. It seems then that the result of operating such a tool will, at best, allow counter-terrorism organisations to focus their existing surveillance and de-radicalisation efforts on a more statistically accurate group.

Creating and Operating a Viable AI Tool for Individual Attacks

In conclusion, the question of what it will take to operate a future predictive counter-terrorism AI system that has some ability to predict possible attackers, as well as how will it work, brings a mixed answer. First, like many other counter-terrorism strategies, it seems that operating such a tool will be much easier in a non-democratic situation. The training set for the AI should include all the possible signals that can be technically collected from previous cases of attacks while filtering them as much as possible to be unbiased and accurate. Operating such a functional machine in liberal democracies will require a dramatic change in the current definitions of the right to privacy as the population should agree to constant monitoring of all the available signals that are created daily by them. The tool recommendations should be as explainable as possible, and it will allow counter-terror agencies to focus their efforts on more statistically accurate

distinct potential attackers whilst applying, as much as possible, soft de-radicalisation techniques to reduce their violent potential which could reduce the number of potential attacks.

The Possible Implications of Using a Theoretically Effective Predictive Counter-Terrorism AI Tool

In the last paragraphs, I focused on the conditions that are needed for the creation and activation of a working predictive counter-terrorism AI tool that can deliver meaningful results. As far as I know, there is still no active tool that follows those conditions (although, once again, similar ones might be already active and hidden from the public) and we can only try to assess what the impact of such a tool will be. The following part will therefore be focused on a theoretical discussion: if an existing preventive tool like that was available, should we use it at all?

Before starting to analyse the tool, it is important once again to present the conclusion this research has arrived at regarding what this kind of tool could achieve and in what conditions. This theoretical tool could statistically increase the pre-identification of potential attackers by suggesting a group of individuals that carry similar signal combinations to previous attackers. For this tool to be effective, it should have constant access to a wide range of signals in the population. The outcome of the recommendation tool will be explainable and collected by the agency that is activating it and it will be handled according to current operational measures such as enhanced surveillance and de-radicalisation strategies.

The Meaning and Pros and Cons of Having a Working Preventive AI Tool Using Lindahl's Model

The current discourse about the benefit of using AI as a predictive tool can be characterised as having two extreme points of view, one is a very positive discussion about the possible benefits of it, and the other is extremely critical of its use, claiming it is both damaging and problematic

(McDaniel and Pease 2021, 2). There are major differences between the two sides. The supporting side presents the possible advantages that will exist when this kind of tool is effective: increasing impact and reducing work for humans (Perrot 2017), more accurate and reducing the need for large scale scanning (Ariel 2019), and reducing bias, which always comes into play in human decisions (Stevenson 2018). The opposing side's criticisms are focused on presenting how the current tools are not effective and how it is impossible to create an effective tool because of bias, legal issues, accuracy and accountability (Babuta 2019; Wall 2024; T. B. Munk 2017; Couchman 2019). No serious discussion has been given to the question of 'if an algorithm can surpass all of those difficulties, would and should we use it?'

In the eyes of law enforcement organisations, counter-terrorism agencies, and some terrorism experts, the answer to that is clear. A predictive AI system that finds the right balance between the clear benefits and potential risks is the future of counter-terrorism (Ganor 2019, 622). That view is understandable as those counter-terrorism organisations usually perceive their prevention goal as just that - finding who the potential attackers are and stopping them before the attack. The prevention can be carried out along with more severe methods such as incarceration and criminal proceedings, and softer methods such as de-radicalisation techniques. Very few counter-terrorism organisations are willing to - or have a mandate to - deal with and ask the more difficult questions like 'what are the root causes for the abrupt outbreak of violence?' and the even more difficult question 'how can we deal with it?' (Zulaika 2009). A clear example of this can be seen in the ISA tool case study. In the article by Barbing and Glick, they present the ISA's analysis of all the pressure points that push young Palestinians to attack such as economic despair, limitations on freedom of movement and lack of support systems (in short,

the Israeli military occupation) (Barbing and Glick 2019), as well as the ISA's conclusion which is to build a technological tool that will gather all of those signals that will tell them who the potential attacker is, instead of thinking of ways to reduce those pressures.

This is the main criticism that Critical Terrorism Studies (CTS) has when it comes to counter-terrorism theories and efforts. Lindahl (2019) sought to present a Weber-style 'ideal type' model of counter-terrorism strategies that incorporate CTS values. This model was developed and used in his book to assess the general counter-terrorism strategies used by countries (in the book, the model was applied to Norwegian counter-terrorism strategies). However, many parts of the model can be used to assess specific counter-terrorism tactics, like a predictive AI tool, as the same logic applies to them. In the following paragraphs, I will examine a theoretical working preventive AI tool, as defined earlier, using Lindahl's critical counter-terrorism framework.

The Lindahl Model - Key assumptions

The first challenge Lindahl's model identifies for any counter-terrorism strategy is dealing with the underlying issues that have led to the outbreak of violence (Lindahl 2019, 92). As the aim of the AI tool is to find similarities between past attacks/attackers and potential new ones, based on the previous statistical patterns of attackers, it does not seem to be very relevant to this stage of the model. I will refer to this stage once again in the conclusion, as I think it could present an interesting challenge for AI developers.

The Lindahl Model - Basic Principles

Lindahl maps a series of basic principles that are the basis of his model.

Dare to know: This principle requires challenging the regular boundaries of counter-terrorism and 're-examining basic concepts, opining up to what has been closed out, re-humanizing what

have been dehumanized… denaturalizing established common sense.' (Lindahl 2019, 93). Whilst the theoretical preventive AI in question does not shutter the ontological walls of counter-terrorism, it does bring a fresh look at counter-terrorism measures, first by looking at many new signals that have not been collected before regarding terrorism – ones that will probably lead to less biased results compared to previous tools. Secondly, its ability to explain its findings could lead to constant feedback about the sources and conditions that lead to humans choosing violence. It probably does not answer Lindahl's radical thinking requirement of 'dare to know', but it might qualify as 'dare to know a little bit'.

The Lindahl Model - Emancipation

The principle of emancipation in Lindahl's model is the idea that we cannot sacrifice the rights of others for our own safety because, by doing so, we will just encourage more violence (Lindahl 2019, 94). In general, the theoretical predictive AI tool challenges the rights of all of the population as it monitors the personal data of everyone without limiting it to a specific group of people. Questions arise when the tool presents its findings and identifies a group of people as potential attackers. The challenge to be faced is whether the enhanced surveillance or de-radicalisation tactics are soft enough not to lead to more violence.

The Lindahl Model Means/Ends Relationship and Non-Violence Principles

In Lindahl's model, these two principles are connected to each other. The means/ends principle says that the change we want to create should determine the means we want to use, as they are interlinked. According to Lindahl, if our goal is to prevent violence, our means can never be violent and this is where the non-violent principle is linked. According to him 'violence is ineffective in achieving both the minimal goal of preventing terrorist attacks, and the maximalist

goal of resolving violent conflicts' (Lindahl 2019, 96). The question of whether using the theoretical AI predictive tool is violent, is not as easy as it sounds. Firstly, many people, including Lindahl himself, see in mass surveillance a violent action because violence has many more layers than just the physical one. Secondly, it returns once again to the question of what will be done with the recommendations made by the tool. On the question of surveillance, it can be argued that it is no longer violent if, like with the theoretical AI tool, the surveillance is done legally, it applies to all of the population, and there is democratic agreement to do so. As for the measures, it is once again a question of what kind of intervention is chosen.

The Lindahl Model - Prevention of Terrorism

The principle of prevention of terrorism, according to Lindahl, does not end in operational success where, for example, certain attacks have been prevented and no successful attacks have taken place over the last year. Lindahl presents three examples of such strategies. One is to strengthen the skills and abilities of the UN in solving and mediating political conflicts. A second is to cut the supply chain of weapons to terrorists by dramatically reducing weapon manufacturing in general. A third is upholding liberal values when countering terrorism and not inventing specific anti-liberal legal mechanisms to fight terror. Lindahl is well aware of the utopian nature of some of these suggestions. However, he insists that aspiring towards them is the only way to break the vicious cycle of violence (Lindahl 2019, 103). It is obvious that, although the theoretical AI tool is a preventive tool, the kind of prevention - statistically identifying potential attackers as they plan an attack - is not the type of prevention tactics that Lindahl presented. However, it can be argued that it is still a prevention thinking tool that, if joined with the right non-violent tactics of intervention, might be a clear improvement on current counter-terrorism strategies.

The Lindahl Model - Evaluation

According to Lindahl, any counter-terrorism strategy or tactic should have a clear means of assessing its effectiveness:

> 'As such, the three tests of effectiveness, proportionality, and legitimacy, all tests of counter terrorism on a specific level, but taken together, they combine for a test of counter terrorism on how many lives have been saved and at what cost financially and ethically' (Lindahl 2019, 105).

Once again, evaluating a preventive strategy that statistically identifies potential attackers will never be accurate, and placing too much trust in theoretical tool predictions can lead to an over optimistic evaluation of how many lives have been saved. An example can be clearly seen in the ISA case study where the heads of the ISA attributed hundreds of prevented attacks to their tool, without very clear signals (Briner 2018). However, a more careful assessment, based on very strong signals such as detailed attack plans, weapon purchases and clear confessions, might help in creating a clearer assessment of effectiveness.

Having said that, this kind of assessment is once again a very important but narrow way to look at counter-terrorism. At this stage, as with any stage of his model, Lindahl asks for a broader look at the impact of using the tool and whether the use of the tool will reduce total violence/suffering. This broader look, which is at the basis of critical terrorism studies is, of course, the weak point of the theoretical predictive AI tool. It is its weak point firstly because it is just a tool and not a wide-ranging strategy. But more than that, it is its weak point because it will be, once again, just another part of the counter-terrorism pendulum (Zulaika 2009).

It is very clear that, even if this theoretical tool worked perfectly, collected data from all of the population and provided a clearer statistical result, it would still create a biased reality. It would be biased but not in the sense that it would look for a specific population because of false data. It would be biased because the reality is many times biased.

Let's take, for example, the Israeli/Palestinian test case. If we apply the theoretical tool there, it will collect signals from all of the population based on all previous acts of violence. It will still predict attackers mostly from two main groups: young Palestinians and young religious settler Jews. The fact that the predictions will be more accurate might reduce the number of successful violent attacks and the impact of counter-terrorism on innocent members of those societies but it certainly will not solve the main political issues that are pushing those groups towards violence. This tool is therefore, even at best, a tool to contain violence and reduce its impact but, in the long-term, it is just another tool of control. The desire of counter-terrorism agencies in developing such a tool is obvious. After all, they often seek the same very short-term work solutions this kind of tool could provide. However, as discussed, the price for creating such an operational tool is not small, both on the operational side with constant analysis of immense layers of data, and on the legal and societal side, allowing a government entity to collect and process massive amounts of our private information.

The Evaluation Result

Lindahl did not create his critical counter terrorism model to score a counter-terrorism strategy or tactic and to conclude a positivistic conclusion about it, good or bad. The model was created as a tool to observe counter-terrorism measures whilst applying critical counter terrorism thinking. Very few, if any, existing counter-terrorism tactics will pass all the elements of the

model, as critical counter-terrorism studies thinking does not usually guide the agenda of counter-terrorism agencies and policy makers. Having said that, academics applying this model to existing counter-terrorism strategies and tactics can help in exposing potential difficulties that are usually unseen by its developers and help suggest a potential alternative. This research concluded that to operate a statistically and legally potentially-effective AI tool that could predict an attacker prior to an attack requires considerable technological abilities which are still not available (for example: a clear explanation of the results of the decision by the AI) and a dramatic change in our understanding of the right to privacy. Applying Lindahl's model to such a theoretical tool is a good way to shed light on the potential issues that can arise from creating and activating it. When measuring the true societal cost of creating and activating this tool against the true possible impact, it seems that the need to develop such a theoretical tool is debatable at best.

Chapter 9 - Conclusions

This research began five years ago with a simple but pressing question: What is the ISA doing with predictive AI? As the research progressed, that initial question evolved into a series of deeper and more refined inquiries. I sought to understand how and why counter-terrorism agencies develop AI tools designed to identify individual attackers. Using the ISA as a case study, I aimed to examine what these agencies believe such tools are capable of, what the tools can actually achieve in practice, whether there is a more responsible or effective way to design and operate them, and what it truly means to deploy such tools when viewed through a critical theoretical lens.

Answering these questions was far from straightforward. As is often the case when studying secretive security institutions, access to data was limited, and transparency was minimal. The conclusions presented here are based solely on the data I was able to obtain, which, while incomplete, still allows for meaningful analysis. Despite these limitations, I believe this research offers a sufficiently clear and nuanced picture of the complex challenges involved in the use of predictive AI technologies in counter-terrorism efforts.

The increasing power of AI's predictive capabilities has naturally drawn the attention of counter-terrorism agencies looking for new ways to combat the rise in so-called "lone wolf" or individual attacks. However, as this research demonstrates through a detailed analysis of the ISA's predictive AI tool, these technologies are deeply problematic. In terms of effectiveness, their short-term ability to identify a specific individual prior to an attack is questionable, and their long-term impact on reducing broader patterns of violence is similarly unconvincing. Beyond their

operational limitations, the societal and ethical costs of such tools are significant and cannot be ignored.

My analysis further reveals that the appeal of predictive AI for agencies like the ISA extends beyond operational efficiency. It also reflects economic and political interests, as well as a problematic tendency to shift moral responsibility from human agents to algorithmic systems—entities that lack the capacity for ethical reasoning or accountability. This moral outsourcing is particularly concerning in the context of counter-terrorism, where life-and-death decisions are at stake.

Recognizing the flaws in the design and deployment of the ISA's actual tool, I turned to a hypothetical question: under what ideal conditions could a predictive AI system be developed that might yield more statistically accurate and ethically sound results? To explore this, I constructed a theoretical model and evaluated it using Lindahl's framework of counter-terrorism. Even under those imagined "perfect" conditions—where the data is clean, the model is unbiased, and legal oversight is strong—the results remained troubling. According to Lindahl's model, the individual and societal costs of operating such a system would still be too high to justify its implementation.

This leads to a critical conclusion: at this stage, there is no ethical, legal, or practical justification for the use of AI tools aimed at predicting who will commit an individual act of violence. Unfortunately, this conclusion does not mean that counter-terrorism agencies will halt the development of such tools. On the contrary, the trend toward predictive technologies is likely to persist, driven by the same political, institutional, and psychological forces explored throughout this study.

This leaves an important and unresolved question: if person-based predictive AI is not a viable solution, what is?

An Ethical Way of Using the Predictive Power of AI to Prevent Incidents of Individual Violence

One of the most common, and sometimes justified, complaints against critical scholars, is that it is very easy to criticise the efforts made by others without suggesting an alternative solution. Incidents of individual violence, no matter how they are framed - as individual terrorist attacks, school shootings, or attacks by individuals with mental health issues, cause terrible losses all over the world and must not be ignored. However, and this is the main argument of critical terrorism studies, the violence should not be addressed in a way that leads to even more violence and more harm. The question I will try to answer in the following paragraphs is whether there is a way we can harness the ever-growing statistical predictive power of AI to prevent violence without creating more violence.

My suggestion, once again following Lindahl's model of critical counter-terrorism, is that the power of predictive AI will be used on a global level, to predict where a trend of individual violence is about to erupt. As we already know, once violence has erupted, and it does not matter if it is between countries, groups or individuals, it is much harder to stop compared to preventing it from erupting in the first place (Malešević 2010). We also know that an outbreak of these waves of violence is very difficult to predict, and that can be the tool's contribution. This tool will be constantly fed with the newest aggregated statistical data from official statistics bodies and will combine this with a constant sentiment semantic analyses of national, regional, and social media open sources. By combining these resources, this tool will produce reports which will statistically

rank which region is in most danger of reaching the conditions for a wave of individual violence. For example, a school shooting in the US is unfortunately a known phenomenon that US agencies are finding very difficult to stop. As Peterson and Densley claim, a unique set of sociological conditions have led to an increase in these tragic events and a tremendous amount of effort is now needed to try and stop them (Peterson and Densley 2022). However, a very similar pattern has just recently begun to appear in Brazil, with school children or ex-students opening fire in their school and killing students and teachers (Roza and Telles 2024). Now think about a tool that could have created a report, based on the statistical data that is collected by each country together with semantic analyses of open media sources. The report could identify the risk of an outbreak of violence in Brazilian schools before it happens, all based on training data already gathered in the US. This output could be used by a team of sociologists and policy makers to try and address the core issues that might be the reason behind such a possible outcome and prevent it. Another potential example, anti-migrant hate, is, at the time this research has been written, gaining momentum all over Europe. It wouldn't be surprising if these sentiments lead to actual violent actions by individuals against migrants or the supporters of migrants. However, the organisations that are trying to battle those sentiments are under resourced and need to choose where to invest their efforts. A tool, like the one described here can, by using the right data, provide a report that ranks the risk according to region, a report that such an organisation could use to prioritise their efforts to de-escalate the situation. By using such a predictive tool to suggest potential social risks that might lead to violence, and not potentially risky individuals, this proposed tool could be both more impactful and much less dangerous.

When examining this kind of potential predictive tool, according to Lindahl's critical terrorism studies model of counter-terrorism (Lindahl 2019), it seems that this theoretical tool fits much better:

In terms of **'Key Assumptions'**, this tool could be used by international bodies in order to identify potential socio-political instabilities that might lead to a wave of violent attacks by individuals, and by already removing itself from any stagnated assumed definition of terrorism or terrorists, those attacks could be framed differently in different regions.

Regarding the **'Basic Principles'**, this tool could try to harness the power of AI to analyse big data and identify statistical connections in social contexts that are invisible to social scientists, and thus it fits the principle of **'Dare to Know'**, as it might provide us with totally new ways of looking at violence and ways to deal with it. As for **'Emancipation'**, the hope is that this tool would not only rank risky regions but also deliver an output on the risk factors that led to the ranking and thus allow those risk factors to be addressed. By doing so, the tool would protect all people - those who are at risk from violence and those who are at risk of turning to violence. Regarding the principles of **'Means/Ends'** relationship and **'Non-violence'**, the system would analyse open-source data that is regularly analysed by policy makers and academics and the only difference would be its ability to deal with the vast amount of data and to find patterns that are missed by human cognition. Its output would be a risk report about areas where such violence might arise so that those areas can get the right resources. Therefore, there is a clear and positive connection between the means and the end result, and no violence will be used at any stage of the tool application. The **'Priorities'** section of the model would also be addressed in full as **'Prevention'** of violence is the tool's only use, and the tool prevention methodology is to give priority to

identifying and then dealing with the core and structural issues that might lead to a rise in violence. The **'Evaluation'** part of the model would be, just as with any preventive tool, not easy to assess. However, we could compare whether there was a reduction in the global amount of incidents of individual violence following the activation of the tool and its predictions, and the response that would be given to the prediction.

It seems that the operation of such a preventive theoretical AI tool fits very well with Lindahl's CTS model of counter-terrorism. It is a non-harmful strategy that could reduce violence by dealing with the core issues. However, as the theoretical framework of Lindahl's model is based on critical thinking, I must also address several major caveats regarding the operation of such a theoretical tool.

First, this tool, even if it worked perfectly, could only predict the possibility of rising tension in a big group of people that might push individuals to act violently. I don't think that this tool, with all its powers, could have identified Norway in 2011 as a country prone to violence prior to the Anders Breivik attack with its disastrous results. Therefore, it should be clear that there should always be a risk of individual violence happening in countries that are not considered at high risk of such attacks. Second, as I have elaborated in the previous chapter, AI tools are, and will be, only as good as the data they are given. National and regional statistical datasets are collected very differently in many places. They are often influenced by political agendas and are not updated in real time (C. W. Howard 2021). This can harm the accuracy of the tool and its ability to identify risk when a swift change is taking place. The third, and the biggest caveat of all, is regarding the actions that should be carried out to reduce the tension that might lead to violence, once such tension is identified by the AI. Socio-political discontent within a group in a particular

region, that could create a rise in individual violence, is often connected to the deep-rooted socio-political problems of that region, and those require not only a lot of resources to solve them but also the cooperation of the socio-political sides involved. Let's assume that the AI tool identifies the city of Belfast as a place where there is a high risk of attacks of individual violence, based on the ongoing political tensions between republicans and loyalists. Does anyone have a good suggestion as to how to solve the core issues underlying this tension that has not been tried in the last 30 years of conflict resolution attempts? Another problem regarding the results of the tool's identification process, is that the de-radicalisation processes of communities are still very questionable and under researched in terms of impact and success (Lindekilde 2016, 255). Some countries might use them, even in good faith, in a way that might create more harm than good, by, for example, singling out 'suspected communities' as being prone to violence (Heath-Kelly 2013).

Even after recognising some of the problems that might be connected with the activation of such tool, this research suggests trying to harness the growing power of predictive AI in order to identify regions of high risk in advance - where a wave of individual violence might start - as well as investing the available resources to deal with the core issues that are the cause of the potential violence. Operating such a tool successfully and without creating more harm requires a global effort and a lot of global good faith. Unfortunately, in the current global political situation, there is very little good faith to be found, so much so that it almost makes this idea a utopian one. However, and with the very clear possibility of overusing a well-known quote by Adriano Olivetti, a failed utopian: 'the term Utopia is often the most convenient way to excuse what you do not have the urge, the ability, or the courage to do'

Bibliography

Abdul-Dayyem, Mariam, and Efrat Ben-Ze'ev. 2020. "The Shahid as a Palestinian Icon: Negotiating Meanings." *British Journal of Middle Eastern Studies* 47 (5): 849–67. https://doi.org/10.1080/13530194.2019.1580184.

Abrams, Elliott, Linda Robinson, Ray Takeyh, and Steven A. Cook. 2024. "One Year After the October 7 Attacks: The Impact on Four Fronts | Council on Foreign Relations." https://www.cfr.org/article/one-year-after-october-7-attacks-impact-four-fronts.

Alfiqra, and A U Khasanah. 2020. "Implementation of Market Basket Analysis Based on Overall Variability of Association Rule (OCVR) on Product Marketing Strategy." *IOP Conference Series. Materials Science and Engineering* (Bristol) 722 (1): 12068. https://doi.org/10.1088/1757-899X/722/1/012068.

Amnesty International. 2022. "Israel's Apartheid against Palestinians." Amnesty International, February 1. https://www.amnesty.org/en/latest/campaigns/2022/02/israels-system-of-apartheid/.

Amnesty International UK. 2018. *UK: Trapped in the Matrix: Secrecy, Stigma, and Bias in the Met's Gangs Database*. Amnesty International UK. https://policehumanrightsresources.org/trapped-in-the-matrix-secrecy-stigma-and-bias-in-the-mets-gangs-database.

Anscombe, F. J. 1973. "Graphs in Statistical Analysis." *The American Statistician* 27 (1): 17–21. https://doi.org/10.1080/00031305.1973.10478966.

Ariel, barak. 2019. "Technology Policing." In *Police Innovation: Contrasting Perspectives*, edited by David Weisbrud. Cambridge University Press. https://www.cambridge.org/core/books/police-innovation/technology-policing/D69D8E9E70A47FB998328758272863EB.

Avis, Maya, Daniel Marciniak, and Maria Sapignoli, eds. 2025. *States of Surveillance: Ethnographies of New Technologies in Policing and Justice*. Routledge Studies in Surveillance. Routledge.

Avraham, Yuval. 2023. "'A Mass Assassination Factory': Inside Israel's Calculated Bombing of Gaza." +972 Magazine, November 30. https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/.

Babuta, Alexander. 2019. "Data Analytics and Algorithmic Bias in Policing." In *Rusi.Org*. https://rusi.org/sites/default/files/20190916_data_analytics_and_algorithmic_bias_in_policing_web.pdf.

Babuta, Alexander. 2020. "Innocent Until Predicted Guilty? Artificial Intelligence and Police Decision-Making." In *RUSI*. https://rusi.org/publication/rusi-newsbrief/innocent-until-predicted-guilty-artificial-intelligence-and-police.

Babuta, Alexander, Marion Oswald, and Ardi Janjeva. 2020. *Artificial Intelligence and UK National Security*.

Babuta, Alexander, Marion Oswald, and Christine Rinik. 2018. *Big Data and Police Decision-Making*. RUSI.

Bachar, Eli. 2020. *Shabak Be-Mivḥan*. Ha-Makhon ha-Yiśre'eli le-demoḳraṭyah.

Backlinko. 2024. "ChatGPT Statistics 2024: How Many People Use ChatGPT?" Backlinko, June 4. https://backlinko.com/chatgpt-stats.

Bandura, Albert. 2004. "The Role of Selective Moral Disengagement in Terrorism and Counterterrorism." In *Understanding Terrorism: Psychosocial Roots, Consequences, and Interventions.*, edited by Fathali M. Moghaddam and Anthony J. Marsella. American Psychological Association. https://doi.org/10.1037/10621-006.

Barbing, Arik, and Or Glick. 2019. "Lone terrorism – the ISA in operation 'Godel Hashaa.'" *Routine security*, The campaign between the wars, no. 23: 127–48.

Barnes, Beau D. 2012. “Confronting the One-Man Wolf Pack: Adapting Law Enforcement and Prosecution Responses to the Threat of Lone Wolf Terrorism Note.” *Boston University Law Review* 92 (5): 1613–62.

Barredo Arrieta, Alejandro, Natalia Díaz-Rodríguez, Javier Del Ser, et al. 2020. “Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI.” *Information Fusion* 58 (June): 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

Basiago, Andrew D. 1994. “The Limits of Technological Optimism.” *Environmentalist* 14 (1): 17–22. https://doi.org/10.1007/BF01902656.

Bastug, Mehmet F., Aziz Douai, and Davut Akca. 2020. “Exploring the ‘Demand Side’ of Online Radicalization: Evidence from the Canadian Context.” *Studies in Conflict & Terrorism* 43 (7): 616–37. https://doi.org/10.1080/1057610X.2018.1494409.

Benbouzid, Bilel. 2019. “To Predict and to Manage. Predictive Policing in the United States.” *Big Data & Society* 6 (1): 2053951719861703. https://doi.org/10.1177/2053951719861703.

Benisho, Nethanel. 2004. “The Criminal Law in Judea Samaria And Gaza : Overture and Trends.” *Mispat Vet’sava* 18: 231.

Ben-Naftali, Orna, Michael Sfard, and Hedi Viterbo. 2019. *The ABC of the OPT*. 1st ed. Cambridge University Press.

Ben-Natan, Smadar. 2014. “The Application of Israeli Law in the Military Courts of the Occupied Palestinian Territory.” *Theory and Criticism* 43 (45).

Benvenisti, Eyal. 1993. “Judicial Misgivings Regarding the Application of International Law: An Analysis of Attitudes of National Courts.” *European Journal of International Law* 4 (2): 159–83. https://doi.org/10.1093/oxfordjournals.ejil.a035824.

Berbing, Erik. 2019. “(Heb). The Lone Terrorist- The GSS in the ‘Size of the Hour’ Campaign.” *Routine Security The Campaign between the Wars* 22–23.

Berda, Yael. 2017a. *Living Emergency: Israel’s Permit Regime in the Occupied West Bank*. Stanford University Press.

Berda, Yael. 2017b. *Living Emergency: Israel’s Permit Regime in the Occupied West Bank*. Stanford University Press.

Bergman, Ronen. 2024. “The ‘secret tool’ of the Amman and the ISA - which led to the neglect of other sources of intelligence (in hebrew).” *Ynet*, September 19. https://www.ynet.co.il/news/article/yokra14072116.

Berntzen, Lars Erik, and Tore Bjørgo. 2021. “The Term ‘Lone Wolf’ and Its Alternatives.” *Perspectives on Terrorism* 15 (3): 132–41. JSTOR.

Blackbourn, Jessie, Nicola McGarrity, and Kent Roach. 2019. “Understanding and Responding to Right Wing Terrorism.” *Journal of Policing, Intelligence and Counter Terrorism* 14 (3): 183–90. https://doi.org/10.1080/18335330.2019.1667014.

Blair, Graeme, Alexander Coppock, and Margaret Moor. 2020. “When to Worry about Sensitivity Bias: A Social Reference Theory and Evidence from 30 Years of List Experiments.” *American Political Science Review* 114 (4): 1297–315. https://doi.org/10.1017/S0003055420000374.

Borger, Julian. 2024. “Israeli Security Chief Condemns ‘Terrorism’ of Militant Settlers.” World News. *The Guardian*, August 23. https://www.theguardian.com/world/article/2024/aug/23/israeli-security-chief-ronen-bar-hilltop-youth-west-bank.

Boyle, Meredith. 2013. “Lone Wolf Terrorism and the Influence of the Internet in France.” *Digital Commons @ Connecticut College*.

Braga, Anthony A, Daniel W Webster, Michael D White, and Hildy Saizow. 2014. *SMART Approaches to Reducing Gun Violence*. Bureau of Justice Assistance.

Brendel, Wieland, and Matthias Bethge. 2019. “Approximating CNNs with Bag-of-Local-Features Models Works Surprisingly Well on ImageNet.” arXiv:1904.00760. Preprint, arXiv, March 20. http://arxiv.org/abs/1904.00760.

Briner, Joshua. 2018. “We Have Foilded 250 Attacks since the Beginning of the Year (in Hebrew).” In *Haaretz*. https://www.haaretz.co.il/news/politics/1.6173611.

Brown, Kathrine, and Elizabeth Pearson. 2018. “Social Media, the Online Enviorment and Terrorisem.” In *Routledge Handbook Of Terrorism And Counterterrorism*, 1st ed., edited by Andrew Silke. Routledge. https://doi.org/10.4324/9781315744636.

B’tselem. 2016. “Presumed Guilty: Remand in Custody by Military Courts in the West Bank | B’Tselem.” http://www.btselem.org/publications/summaries/201506_presumed_guilty.

Carmeli, Lior. 2019. “Operation ‘Godel Hashaa’ 2015 2016 - Learning and developing an operational answer.” *Routine security*, The campaign between the wars, no. 17: 137–48.

Carr, Caleb. 2007. “‘Terrorism’: Why the Definition Must Be Broad.” *World Policy Journal* 24 (1): 47–50. https://doi.org/10.1162/wopj.2007.24.1.47.

Catignani, Sergio. 2017. “The Strategic Impasse in Low-Intensity Conflicts: The Gap Between Israeli Counter-Insurgency Strategy and Tactics During the Al-Aqsa Intifada.” In *Warfare in the Middle East since 1945*. Routledge. https://doi.org/10.4324/9781315234304-19.

Chomsky, Noam. 1979. *The Political Economy of Human Rights. v.2, After the Cataclysm: Postwar Indochina and the Reconstruction of Imperial Ideology / Noam Chomsky and Edward S. Herman.* With Edward S. Herman and Bertrand Russell Peace Foundation. Spokesman.

Cohen, Gerald A. 2009. *If You’re an Egalitarian, How Come You’re So Rich?* Harvard University Press.

Cohen, Stuart Alan, and Aharon S. Klieman. 2019. *Routledge Handbook on Israeli Security*. Routledge Handbooks. Routledge.

Conway, Maura. 2016. “Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research.” *Studies in Conflict & Terrorism* 40 (1): 77–98. https://doi.org/10.1080/1057610x.2016.1157408.

Conway, Maura. 2017. “Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research.” *Studies in Conflict & Terrorism* 40 (1): 77–98. https://doi.org/10.1080/1057610X.2016.1157408.

Cornish, Paul. 2010. “Technology, Strategy and Counterterrorism.” *International Affairs* 86 (4): 875–88. https://doi.org/10.1111/j.1468-2346.2010.00917.x.

Couchman, Hanna. 2019. *Report: Policing by Machine*. Liberty. https://www.libertyhumanrights.org.uk/issue/policing-by-machine/.

Dajani Daoudi and Barakat. 2013. “Israelis and Palestinians: Contested Narratives.” *Israel Studies* 18 (2): 53. https://doi.org/10.2979/israelstudies.18.2.53.

Daniel Kasbari, Carol. 2022. “Once Again, Israel Throws up an Unlawful Barrier to Palestinian Family Reunification.” Middle East Institute, March 15. https://www.mei.edu/publications/once-again-israel-throws-unlawful-barrier-palestinian-family-reunification.

DG. 2005. *Terrorist Recruitment: Addressing the Factors Contributing to Violent Radicalisation*. https://primarysources.brillonline.com/browse/human-rights-documents-online/communication-from-the-commission-to-the-european-parliament-and-the-council;hrdhrd46790058.

Dinstein, Yoram. 2009. *The International Law of Belligerent Occupation*. Cambridge University Press.

Doaa Abu Elyounes. 2020. “Bail or Jail?” *Science and Technology Law Review*, July 29, 376-445 Pages. 376-445 Pages. https://doi.org/10.7916/STLR.V21I2.6838.

Dorfman, Yossi. 2022. “Death to arabs is today Death to terrorists.” *The hotest place in hell*, July 26. https://www.ha-makom.co.il/post-yosi-the-new-bengvir/.

Eiran, Ehud, and Peter Krause. 2018. “Old (Molotov) Cocktails in New Bottles? ‘Price-Tag’ and Settler Violence in Israel and the West Bank.” *Terrorism and Political Violence* 30 (4): 637–57. https://doi.org/10.1080/09546553.2016.1194271.

Elshimi, M. S. 2017. *De-Radicalisation in the UK Prevent Strategy: Security, Identity, and Religion*. Routledge Critical Terrorism Studies. Routledge/Taylor & Francis Group.

Englund, Scott, and Michael Stohl. 2016. “Constructions of Terrorism.” *Perspectives on Terrorism* 10 (3): 33–39.

Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. First Edition. St. Martin’s Press.

Evans, Sarah Austin and Jonathan. 2023. “Israelis Have Grown More Skeptical of a Two-State Solution.” *Pew Research Center*, September 26. https://www.pewresearch.org/short-reads/2023/09/26/israelis-have-grown-more-skeptical-of-a-two-state-solution/.

Feier, Till, Jan Gogoll, and Matthias Uhl. 2021. “Hiding Behind Machines: When Blame Is Shifted to Artificial Agents.” *Papers*, Papers, January, 2101.11465.

Ferguson, Andrew. 2012. “Predictive Policing and Reasonable Suspicion.” *Emory Law Journal* 62 (2): 259.

Ferguson, Andrew. 2017. “Policing Predictive Policing.” *Washington University Law Review* 94 (5): 1109–89.

Fernandez, Miriam, and Harith Alani. 2021. “Artificial Intelligence and Online Extremism: Challenges and Opportunities.” In *Predictive Policing and Artificial Intelligence*, edited by John McDaniel Pease Ken. Routledge. https://doi.org/10.4324/9780429265365.

Fitzgerals, James. 2016. “Critical Epistimlogies of Terrorism.” In *Routledge Handbook of Critical Terrorism Studies / Edited by Richard Jackson.*, edited by Richard Jackson. Routledge.

Floridi, Luciano. 2024. “Why the AI Hype Is Another Tech Bubble.” *Philosophy & Technology* 37 (4): 128. https://doi.org/10.1007/s13347-024-00817-w.

Floridi, Luciano, and J. W. Sanders. 2001. *Artificial Evil and the Foundation of Computer Ethics*. Edited by Luciano Floridi and J. W. Sanders. Springer Netherlands. https://philarchive.org/rec/FLOAEA.

Flyvbjerg, Bent. 2006. “Five Misunderstandings About Case-Study Research.” *Qualitative Inquiry* 12 (2): 219–45. https://doi.org/10.1177/1077800405284363.

Fritz, Alexis, Wiebke Brandt, Henner Gimpel, and Sarah Bayer. 2020. “Moral Agency without Responsibility? Analysis of Three Ethical Models of Human-Computer Interaction in Times of Artificial Intelligence (AI).” *De Ethica* 6 (1): 1. https://doi.org/10.3384/de-ethica.2001-8819.20613.

Frostenson, Sarah. 2015. “Most Terrorist Attacks in the US Are Committed by Americans — Not Foreigners.” Vox, November 23. https://www.vox.com/2015/11/23/9765718/domestic-terrorism-threat.

Fuchs, Andrew, Andrea Passarella, and Marco Conti. 2023. “Modeling, Replicating, and Predicting Human Behavior: A Survey.” *ACM Transactions on Autonomous and Adaptive Systems* 18 (2): 1–47. https://doi.org/10.1145/3580492.

Ganor, Boaz. 2019. “Artificial or Human: A New Era of Counterterrorism Intelligence?” *Studies in Conflict & Terrorism*, 1–20. https://doi.org/10.1080/1057610x.2019.1568815.

Ganor, Boaz. 2021. “Understanding the Motivations of ‘Lone Wolf’ Terrorists: The ‘Bathtub’ Model.” *Perspectives on Terrorism* 15 (2): 23–32.

Gil de Zúñiga, Homero, Manuel Goyanes, and Timilehin Durotoye. 2024. “A Scholarly Definition of Artificial Intelligence (AI): Advancing AI as a Conceptual Framework in Communication Research.” *Political Communication* 41 (2): 317–34. https://doi.org/10.1080/10584609.2023.2290497.

Gill, Paul, Emily Corner, Maura Conway, Amy Thornton, Mia Bloom, and John Horgan. 2017. “Terrorist Use of the Internet by the Numbers.” *Criminology & Public Policy* 16 (1): 99–117. https://doi.org/10.1111/1745-9133.12249.

Gill, Paul, John Horgan, and Paige Deckert. 2014. “Bombing Alone: Tracing the Motivations and Antecedent Behaviors of Lone-Actor Terrorists,,.” *Journal of Forensic Sciences* 59 (2): 425–35. https://doi.org/10.1111/1556-4029.12312.

Giovagnoli, Raffaela. 2013. “‘Computational Ontology and Deontology.’” In *Philosophy and Theory of Artificial Intelligence*, edited by Vincent C. Müller. Studies in Applied Philosophy, Epistemology and Rational Ethics. Springer. https://doi.org/10.1007/978-3-642-31674-6_13.

Goffman, Roman. 2019. “‘From ongoing security to governance’ – a journey of change and scholarship at operation ‘Godel Hashaa’.” *Routine security*, The campaign between the wars, no. 17: 149.

Golan, Daphna. 1990. *The Military Court System in the West Bank - Follow up Report*. B’tselem.

Gonella, F., C. M. V. B. Almeida, G. Fiorentino, et al. 2019. “Is Technology Optimism Justified? A Discussion towards a Comprehensive Narrative.” *Journal of Cleaner Production* 223 (June): 456–65. https://doi.org/10.1016/j.jclepro.2019.03.126.

Greenwood, Shannon. 2022. “3. Internet, Smartphone and Social Media Use.” *Pew Research Center’s Global Attitudes Project*, December 6. https://www.pewresearch.org/global/2022/12/06/internet-smartphone-and-social-media-use-in-advanced-economies-2022/.

Gross, Aeyal M. 2017. *The Writing on the Wall: Rethinking the International Law of Occupation*. Cambridge University Press.

Gruenewald, Jeff, Steven Chermak, and Joshua D. Freilich. 2013. “Overview of: ‘Distinguishing “Loner” Attacks from Other Domestic Extremist Violence: A Comparison of Far-Right Homicide Incident and Offender Characteristics.’” *Criminology & Public Policy* 12 (1): 63–64. https://doi.org/10.1111/1745-9133.12009.

Guay, Jean-Pierre, and Geneviève Parent. 2018. “Broken Legs, Clinical Overrides, and Recidivism Risk: An Analysis of Decisions to Adjust Risk Levels With the LS/CMI.” *Criminal Justice and Behavior* 45 (1): 82–100. https://doi.org/10.1177/0093854817719482.

Gundhus, Helene O. I., and Christin Thea Wathne. 2024. “Resistance to Platformization: Palantir in the Norwegian Police.” *Information, Communication & Society* 27 (13): 2381–99. https://doi.org/10.1080/1369118X.2024.2325533.

Hajjar, Lisa. 2005. *Courting Conflict*. University of California Press.

Hamid, Nafees, and Cristina Ariza. 2022. *Offline Versus Online Radicalisation: Which Is the Bigger Threat?* https://gnet-research.org/2022/02/21/offline-versus-online-radicalisation-which-is-the-bigger-threat/.

Harel, Amos. 2019. “How Israel Stopped a Third Palestinian Intifada.” *Haaretz*, October 4. https://www.haaretz.com/israel-news/2019-10-04/ty-article/.premium/how-israel-stopped-a-third-palestinian-intifada/0000017f-e355-df7c-a5ff-e37f99d30000?v=1655910039370.

Harkov, Lahav. 2015. “Bennett to Europe: Israel Is on the Front Lines in War against Terror and Radical Islam - The Jerusalem Post.” https://www.jpost.com/israel-elections/bennett-to-europe-israel-is-on-the-front-lines-in-war-against-terror-and-radical-islam-391333.

Hasson, Nir. 2014. “The Dangerous, Unpredictable Anomaly of East Jerusalem - Haaretz Com - Haaretz.Com.” https://www.haaretz.com/2014-10-31/ty-article/.premium/the-dangerous-anomaly-of-east-jerusalem/0000017f-f109-d8a1-a5ff-f18b00670000.

Heath-Kelly, Charlotte. 2013. “Counter-Terrorism and the Counterfactual: Producing the ‘Radicalisation’ Discourse and the UK PREVENT Strategy.” *The British Journal of Politics and International Relations* 15 (3): 394–415. https://doi.org/10.1111/j.1467-856X.2011.00489.x.

Herman, Edward S., and Gerry O'Sullivan. 1989. *The "Terrorism" Industry: The Experts and Institutions That Shape Our View of Terror*. 1st ed. Pantheon Books.
Herzog, Lisa. 2018. *Reclaiming the System: Moral Responsibility, Divided Labour, and the Role of Organizations in Society*. First edition. Oxford University Press.
Hever, Shir. 2018. "Securing the Occupation in East Jerusalem: Divisions in Israeli Policy." *Jerusalem Quaterly* 75: 104.
Howard, Cosmo Wyndham. 2021. *Government Statistical Agencies and the Politics of Credibility*. Cambridge Studies in Comparative Public Policy. Cambridge University Press. https://doi.org/10.1017/9781108867962.
Howard, Nigel. 2024. "Amidst Complex Threats, How Can the EU Fight Terrorism More Effectively?" Centre for European Reform. https://www.cer.eu/insights/amidst-complex-threats-how-can-eu-fight-terrorism-more-effectively.
Human Rights Watch. 2017. *China: Police 'Big Data' Systems Violate Privacy, Target Dissent*. November 19. https://www.hrw.org/news/2017/11/19/china-police-big-data-systems-violate-privacy-target-dissent.
Hung, Tzu-Wei, and Chun-Ping Yen. 2021. "On the Person-Based Predictive Policing of AI." *Ethics and Information Technology* 23 (3): 165–76. https://doi.org/10.1007/s10676-020-09539-x.
Hunt, Paul. 1987. *Justice? The Military Court System in the Israeli Occupied Territories*. Al-HAQ.
IEP. 2025. *2024 Global Terrorism Index*. https://www.economicsandpeace.org/wp-content/uploads/2024/02/GTI-2024-web-290224.pdf.
Iliadis, Andrew, and Amelia Acker. 2022. "The Seer and the Seen: Surveying Palantir's Surveillance Platform." *The Information Society* 38 (5): 334–63. https://doi.org/10.1080/01972243.2022.2100851.
Jackson, Richard. 2005. *Writing the War on Terrorism: Language, Politics, and Counter-Terrorism*. New Approaches to Conflict Analysis. Manchester University Press ; Distributed exclusively in the USA by Palgrave.
Jackson, Richard. 2007. "The Core Commitments of Critical Terrorism Studies." *European Political Science* 6: 244–51. https://doi.org/10.1057/palgrave.eps.2210141.
Jackson, Richard. 2016. *Routledge Handbook of Critical Terrorism Studies / Edited by Richard Jackson.* Routledge.
Jackson, Richard, Marie Breen Smyth, and Jeroen Gunning, eds. 2009. *Critical Terrorism Studies: A New Research Agenda*. 1. publ. Critical Terrorism Studies. Routledge.
Jackson, Richard, Lee Jarvis, Marie Breen Smyth, and Jeroen Gunning, eds. 2011. *Terrorism: A Critical Introduction*. Palgrave Macmillan.
Jackson, Richard, Marie Smith, and Jeroen Gunning. 2009. "Critical Terrorism Studies Framing a New Research Agenda." In *Critical Terrorism Studies: A New Research Agenda*, 1. publ, edited by Richard Jackson. Critical Terrorism Studies. Routledge.
Jarvis, Lee. 2009. "The Spaces and Faces of Critical Terrorism Studies." *Security Dialogue* 40 (1): 5–27. https://doi.org/10.1177/0967010608100845.
Jarvis, Lee. 2016. "Critical Terrorism Studies after 9/11." In *Routledge Handbook of Critical Terrorism Studies*, edited by Richard Jackson. Routledge.
Jensen, Benjamin M, Christopher Whyte, and Scott Cuomo. 2020. "Algorithms at War: The Promise, Peril, and Limits of Artificial Intelligence." *International Studies Review* 22 (3): 526–50. https://doi.org/10.1093/isr/viz025.
Johnson, Deborah G., and Mario Verdicchio. 2019. "AI, Agency and Responsibility: The VW Fraud Case and Beyond." *AI & SOCIETY* 34 (3): 639–47. https://doi.org/10.1007/s00146-017-0781-9.

Jore, Sissel H. 2021. “Ontological and Epistemological Challenges of Measuring the Effectiveness of Urban Counterterrorism Measures.” *Security Journal* 34 (2): 231–46. https://doi.org/10.1057/s41284-019-00221-6.

Judea, The military courts in and Sumeria. 2016. *Yearly Statistics Reports- 2007- 2016*. IDF.

Kahneman, Daniel. 2012. *Thinking, Fast and Slow*. Penguin Psychology. Penguin Books.

Kampf, Zohar. 2012. “From ‘There Are No Palestinian People’ to ‘Sorry for Their Suffering’: Israeli Discourse of Recognition of the Palestinians.” *Journal of Language and Politics* 11 (3): 427–47. https://doi.org/10.1075/jlp.11.3.06kam.

Kane, Alex. 2016. “Israel Targeting Palestinian Protesters on Facebook.” The Intercept, July 7. https://theintercept.com/2016/07/07/israel-targeting-palestinian-protesters-on-facebook/.

Keller, Matthew C., Barbara L. Fredrickson, Oscar Ybarra, et al. 2005. “A Warm Heart and a Clear Head: The Contingent Effects of Weather on Mood and Cognition.” *Psychological Science* 16 (9): 724–31. https://doi.org/10.1111/j.1467-9280.2005.01602.x.

Kennedy, Leslie W., Joel M. Caplan, and Eric Piza. 2011. “Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies.” *Journal of Quantitative Criminology* 27 (3): 339–62. https://doi.org/10.1007/s10940-010-9126-2.

Khan, Alamgir, and Christian Kaunert. 2023. “US Drone Strikes, Securitization Processes and Practices: A Case Study of Pakistan.” *Critical Studies on Terrorism* 16 (2): 287–304. https://doi.org/10.1080/17539153.2023.2179571.

Kingsley, Patrick, and Ronen Bergman. 2021. “Israel’s Military Inflicted a Heavy Toll. But Did It Achieve Its Aim?” World. *The New York Times*, May 21. https://www.nytimes.com/2021/05/21/world/middleeast/israel-gaza-war-ceasefire.html.

Krebs, Shiri. 2012. “Lifting the Veil of Secrecy: Judicial Review of Administrative Detentions in the Israeli Supreme Court.” *Vanderbilt Journal of Transnational Law* 45: 639.

Kretzmer, David. 2012. “The Law of Belligerent Occupation in the Supreme Court of Israel.” *International Review of the Red Cross* 94 (885): 207–36. https://doi.org/10.1017/s1816383112000446.

Kruglanski, Arie W., Xiaoyan Chen, Mark Dechesne, Shira Fishman, and Edward Orehek. 2009. “Fully Committed: Suicide Bombers’ Motivation and the Quest for Personal Significance.” *Political Psychology* 30 (3): 331–57. https://doi.org/10.1111/j.1467-9221.2009.00698.x.

Kruglanski, Arie W., Michele J. Gelfand, Jocelyn J. Bélanger, Anna Sheveland, Malkanthi Hetiarachchi, and Rohan Gunaratna. 2014. “The Psychology of Radicalization and Deradicalization: How Significance Quest Impacts Violent Extremism.” *Political Psychology* 35: 69–93.

Kundnani, Arun. 2012. “Radicalisation: The Journey of a Concept.” *Race & Class* 54 (2): 3–25. https://doi.org/10.1177/0306396812454984.

Lalljee, Mansur, Margaret Watson, and Peter White. 1982. “Explanations, Attributions and the Social Context of Unexpected Behaviour.” *European Journal of Social Psychology* 12 (1): 17–29. https://doi.org/10.1002/ejsp.2420120102.

Lekach, Zvi. 2017. “Human Rights in the Military Courts.” In *Mordehai Karmnitzers Book*, 1st ed., edited by Ariel Bendor. Nevo.

Lettau, Jacqueline. 2021. “The Impact of Children’s Academic Competencies and School Grades on Their Life Satisfaction: What Really Matters?” *Child Indicators Research* 14 (6): 2171–95. https://doi.org/10.1007/s12187-021-09830-3.

Levano, Jessie. 2024. “Predictive Policing in the AI Act: Meaningful Ban or Paper Tiger?” *European Law Blog*, ahead of print, July 5. https://doi.org/10.21428/9885764c.6d0aa28c.

Lindahl, Sondre. 2019. *A Critical Theory of Counterterrorism: Ontology, Epistemology and Normativity*. First issued in paperback. Routledge Critical Terrorism Studies. Routledge, Taylor & Francis Group.

Lindekilde, Lasse. 2016. “Radicalization, De-Radicalization, and Counter-Radicalization.” In *Routledge Handbook of Critical Terrorism Studies / Edited by Richard Jackson.*, edited by Richard Jackson. Routledge.

Lindekilde, Lasse, Stefan Malthaner, and Francis O’Connor. 2019. “Peripheral and Embedded: Relational Patterns of Lone-Actor Terrorist Radicalization.” *Dynamics of Asymmetric Conflict* 12 (1): 20–41. https://doi.org/10.1080/17467586.2018.1551557.

Loadenthal, Michael. 2014. “Reproducing a Culture of Martyrdom.” In *Motherhood and War: International Perspectives*, edited by Dana Cooper and Claire Phelan. Palgrave Macmillan US. https://doi.org/10.1057/9781137437945_11.

Loewenstein, Antony. 2024. *The Palestine Laboratory: How Israel Exports the Technology of Occupation Around the World*. Verso Books.

Luitse, Dieuwertje, and Wiebke Denkena. 2021. “The Great Transformer: Examining the Role of Large Language Models in the Political Economy of AI.” *Big Data & Society*, ahead of print, September 29. Sage UK: London, England. https://doi.org/10.1177/20539517211047734.

Lygre, Ragnhild B., Jarle Eid, Gerry Larsson, and Magnus Ranstorp. 2011. “Terrorism as a Process: A Critical Review of Moghaddam’s ‘Staircase to Terrorism.’” *Scandinavian Journal of Psychology* 52 (6): 609–16. https://doi.org/10.1111/j.1467-9450.2011.00918.x.

Mahon, Anastassiya. 2022. “Counter-Terrorism Laws and Freedom of Expression: Global Perspectives: Edited by Téwodros Workneh and Paul Haridakis, Lexington Books, 2021, Lanham, Maryland, USA, and London, UK, 408 Pp., $125.00 (£96.00), Hardback, ISBN 9781793622167.” *Critical Studies on Terrorism* 15 (3): 757–58. https://doi.org/10.1080/17539153.2022.2093885.

Malešević, Siniša. 2010. *The Sociology of War and Violence*. Cambridge University Press.

Mann, Itamar, and Omer Shatz. 2010. “The Necessity Procedure: Laws of Torture in Israel and Beyond, 1987 - 2009.” *Student Scholarship Papers*, January 1. https://openyls.law.yale.edu/handle/20.500.13051/5624.

Martini, Alice, ed. 2021. *Bringing Normativity into Critical Terrorism Studies*. Routledge Critical Terrorism Studies. Routledge.

Mayson, Sandra G. 2018. “Bias In, Bias Out.” SSRN Scholarly Paper 3257004. Rochester, NY, September 28. https://papers.ssrn.com/abstract=3257004.

McCauley, Clark, and Sophia Moskalenko. 2008. “Mechanisms of Political Radicalization: Pathways Toward Terrorism.” *Terrorism and Political Violence* 20 (3): 415–33. https://doi.org/10.1080/09546550802073367.

McDaniel, John L. M., and K. Pease, eds. 2021. *Predictive Policing and Artificial Intelligence*. Routledge Frontiers of Criminal Justice. Routledge.

McKendrick, Kathleen. 2019. “Artificial Intelligence Prediction and Counterterrorism.” In *Chatham House*. https://www.chathamhouse.org/publication/artificial-intelligence-prediction-and-counterterrorism.

McKernan, Bethan, and Harry Davies. 2024. “‘The Machine Did It Coldly’: Israel Used AI to Identify 37,000 Hamas Targets.” World News. *The Guardian*, April 3. https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes.

McQuade, Joseph. 2020. *A Genealogy of Terrorism: Colonial Law and the Origins of an Idea*. Cambridge University Press. https://doi.org/10.1017/9781108896238.

Meehl, Paul E. (Paul Everett). 1954. *Clinical versus Statistical Prediction : A Theoretical Analysis and a Review of the Evidence / Paul E. Meehl.* In *CLINICAL VERSUS STATISTICAL PREDICTION*. University of Minnesota Press.

Metz, Cade. 2016. “In Two Moves, AlphaGo and Lee Sedol Redefined the Future.” Tags. *Wired*. https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/.

Moghaddam, Fathali M. 2005. “The Staircase to Terrorism: A Psychological Exploration.” *American Psychologist* 60 (2): 161–69. https://doi.org/10.1037/0003-066X.60.2.161.
Monaghan, Jeffrey, and Adam Molnar. 2016. “Radicalisation Theories, Policing Practices, and ‘the Future of Terrorism?’” *Critical Studies on Terrorism* 9 (3): 393–413. https://doi.org/10.1080/17539153.2016.1178485.
Monahan, John. 2012. “The Individual Risk Assessment of Terrorism.” *Psychology, Public Policy, and Law* 18 (2): 167–205. https://doi.org/10.1037/a0025792.
Mor, Liron. 2019. “Resistance into Incitement: Translation, Legislation, ‘Early Detection,’ and the Palestinian Poet’s Intention.” *The Arab Studies Journal* 27 (1): 118–55.
Müller, Vincent C., ed. 2018. *Philosophy and Theory of Artificial Intelligence 2017*. Vol. 44. Studies in Applied Philosophy, Epistemology and Rational Ethics. Springer International Publishing. https://doi.org/10.1007/978-3-319-96448-5.
Munk, Timme. 2019. “100,000 False Positives for Every Real Terrorist: Why Anti-Terror Algorithms Don’t Work.” In *Journals.Uic.Edu*. https://journals.uic.edu/ojs/index.php/fm/article/view/7126/6522.
Munk, Timme Bisgaard. 2017. “100,000 False Positives for Every Real Terrorist: Why Anti-Terror Algorithms Don’t Work.” *First Monday*, ahead of print. https://doi.org/10.5210/fm.v22i9.7126.
Nasser-Eddine, Minerva, and Gilbert Caluya. 2011. *Countering Violent Extremism (CVE) Literature Review*. DEFENSE TECHNICAL INFORMATION CENTER. https://apps.dtic.mil/sti/citations/ADA543686.
Neta Ziv. 2018. “Navigating the Judicial Terrain Under Israeli Occupation: Palestinian and Israeli Lawyers in the Military Courts.” *Fordham International Law Journal* 42 (2): 729.
Neumann, Peter R. 2009. *Old and New Terrorism: Late Modernity, Globalization and the Transformation of Political Violence*. Polity Press.
Neumann, Peter R. 2013. “Options and Strategies for Countering Online Radicalization in the United States.” *Studies in Conflict & Terrorism* 36 (6): 431–59. https://doi.org/10.1080/1057610X.2013.784568.
Nilsson, Nils J. 2009. *The Quest for Artificial Intelligence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819346.
Numa, Roni, and Rom Liraz. 2019. “‘Winning and staying human’ The challenges of the central region command at operation ‘Godel Hashaa’.” *Routine security*, The campaign between the wars, no. 17: 123–36.
Ojanen, Tuomas. 2010. “Terrorist Profiling: Human Rights Concerns.” *Critical Studies on Terrorism* 3 (2): 295–312. https://doi.org/10.1080/17539153.2010.491343.
Pantucci, Raffaello. 2011. “What Have We Learned about Lone Wolves from Anders Behring Breivik?” *Perspectives on Terrorism* 5 (5/6): 27–42.
Pappe, Ilan. 2009. “De-Terrorising the Palestinian National Struggle: The Roadmap to Peace.” *Critical Studies on Terrorism* 2 (2): 127–46. https://doi.org/10.1080/17539150903021399.
Pedahzur, Ami. 2009. *The Israeli Secret Services and the Struggle Against Terrorism*. Columbia University Press.
Perri, Jacob. 1999. *Haba Lehargecha*. With יוסי. קשת.
Perrot, Patrick. 2017. “What about AI in Criminal Intelligence? From Predictive Policing to AI Perspectives.” *European Law Enforcement Research Bulletin*, no. 16 (August): 16.
Peterson, Jillian, and James Densley. 2022. *Violence Project How to Stop a Mass Shooting Epidemic*. https://www.abramsbooks.com/product/violence-project_9781419752964/.
Pettinger, Tom. 2021. “CTS and Normativity: The Essentials of Preemptive Counter-Terrorism Intervantions.” In *Bringing Normativity into Critical Terrorism Studies*, edited by Alice Martini. Routledge Critical Terrorism Studies. Routledge.

Phillips, Peter J., and Gabriela Pohl. 2012. “Economic Profiling of the Lone Wolf Terrorist: Can Economics Provide Behavioral Investigative Advice?” *Journal of Applied Security Research* 7 (2): 151–77. https://doi.org/10.1080/19361610.2012.656250.

Puar, Jasbir K., and Amit S. Rai. 2002. “Monster, Terrorist, Fag: The War on Terrorism and the Production of Docile Patriots.” *Social Text* 20 (3 (72)): 117–48. https://doi.org/10.1215/01642472-20-3_72-117.

Qiang, Xiao. 2019. “President XI’s Surveillance State.” *Journal of Democracy* 30 (1): 53–67. https://doi.org/10.1353/jod.2019.0004.

Ramati, Nery. 2019. “The Rulings of the Israeli Military Courts and International Law.” *Journal of Conflict and Security Law* 25 (1): 149–69. https://doi.org/10.1093/jcsl/krz017.

Ramati, Nery, and Karin Torn Hibler. 2021. “The cooperation betwean the police and the Israeli Security Agency in investigating securty offenses.” In *Law and Policing*, edited by Nomi Levenkron and Tamar Kricheli-Katch. Tel Aviv University.

Rawls, John. 1993. *Political Liberalism*. Vol. 177. Columbia University Press.

Reuters. 2020. “NSA Surveillance Exposed by Snowden Was Illegal, Court Rules Seven Years On.” US News. *The Guardian*, September 3. https://www.theguardian.com/us-news/2020/sep/03/edward-snowden-nsa-surveillance-guardian-court-rules.

Ronen, David. 1989. *Shenat Shabak: Ha-He‘arkhut Bi-Yehudah Ṿe-Shomron, Shanah Rishonah*. Miśrad ha-biṭaḥon.

Roza, Thiago Henrique, and Lisieux Elaine de Borba Telles. 2024. “The Rise of School Shootings and Other Related Attacks in Brazil.” *The Lancet Regional Health – Americas* 33 (May). https://doi.org/10.1016/j.lana.2024.100724.

Saban, Itizik. 2018. “The ISA Has Thwarted 250 Attacks since the Beginning of the Year (in Hebrew).” In *Israelhayom.Co.Il*. https://www.israelhayom.co.il/article/563493.

Sageman, Marc. 2008. *Leaderless Jihad: Terror Networks in the Twenty-First Century*. University of Pennsylvania Press.

Sageman, Marc. 2009. “The Next Generation of Terror – Foreign Policy.” *Foreign Policy*, August 10, 36–42.

Salatino, Adriana, Arthur Prével, Emilie Caspar, and Salvatore Lo Bue. 2025. “Influence of AI Behavior on Human Moral Decisions, Agency, and Responsibility.” *Scientific Reports* 15 (1): 12329. https://doi.org/10.1038/s41598-025-95587-6.

Scheffler, Samuel, and Véronique Munoz-Dardé. 2005. “The Division of Moral Labour.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 79: 229–84. JSTOR.

Schmid, Alex. 2013. “Radicalisation, De-Radicalisation, Counter-Radicalisation: A Conceptual Discussion and Literature Review.” *Terrorism and Counter-Terrorism Studies*, ahead of print. https://doi.org/10.19165/2013.1.02.

Schröter, Marie. 2020. *Artificial Intelligence and Countering Violent Extremism: A Primer*. https://gnet-research.org/2020/09/28/artificial-intelligence-and-countering-violent-extremism-a-primer/.

Schuurman, Bart, Edwin Bakker, Paul Gill, and Noémie Bouhana. 2018. “Lone Actor Terrorist Attack Planning and Preparation: A Data-Driven Analysis,.” *Journal of Forensic Sciences* 63 (4): 1191–200. https://doi.org/10.1111/1556-4029.13676.

Scrivens, Ryan, and Garth Davies. 2018. “Identifying Radical Content Online.” Policy Options. https://policyoptions.irpp.org/magazines/january-2018/identifying-radical-content-online/.

Shahaf, Tal. 2020. “ISA cyber chief: ‘From looking on 70 likes on Facebook, I know more about you than you know about yourself.’” YNET, November 27. https://www.ynet.co.il/articles/0,7340,L-5851279,00.html.

Shamgar, Meir. 1982. *Military Government in the Territories Administered by Israel 1967-1980*. Hebrew Univ. Jerusalem Faculty of Law The Harry Sacher Institute for Legislative Research and Comparative Law.
Shanahan, Timothy. 2016. "The Definition of Terrorism." In *Routledge Handbook of Critical Terrorism Studies / Edited by Richard Jackson.*, edited by Richard Jackson. Routledge.
Sharon, Jeremy. 2022. "NGO Says Only 6% of Police Probes of Settler Violence It Was Party to Ended in Charges." https://www.timesofisrael.com/ngo-says-only-6-of-police-probes-of-settler-violence-it-was-party-to-ended-in-charges/.
Shehadeh, Raja. 1988. "Occupier's Law and the Uprising." *Journal of Palestine Studies* 17 (3): 24–37. https://doi.org/10.1525/jps.1988.17.3.00p0004a.
Shehadeh, Raja. 1997. *From Occupation to Interim Accords: Israel and the Palestinian Territories*. In *From Occupation to Interim Accords: Israel and the Palestinian Territories*. Brill. https://brill.com/view/title/14416.
Shinar, Adam. 2001. "Reflections on the Judgments Regarding." *Israel Law Review* 35 (1): 153–68. https://doi.org/10.1017/S0021223700012139.
Shragai, Nadav. 2019. "Finding Terrorist Needles in the Internet Haystack." Magazine. *Www.Israelhayom.Com*, November 1. https://www.israelhayom.com/2019/11/01/finding-terrorist-needles-in-the-internet-haystack/.
Sikimic, Simona. 2015. "Netanyahu: Palestinian Incitement, Not Settlements behind Violence." Middle East Eye, October 16. https://www.middleeasteye.net/news/netanyahu-palestinian-incitement-not-settlements-behind-violence.
Silke, Andrew. 2009. "Contemporary Terrorism Studies Issues in Research." In *Critical Terrorism Studies: A New Research Agenda*, 1. publ, edited by Richard Jackson. Critical Terrorism Studies. Routledge.
Silke, Andrew, ed. 2019. *Routledge Handbook of Terrorism and Counterterrorism*. Routledge Handbooks. Routledge.
Simon, Jeffrey D. 2016. *Lone Wolf Terrorism: Understanding the Growing Threat*. Prometheus Books.
Smith, Brent L., Jeff Gruenewald, Paxton Roberts, and Kelly R. Damphousse. 2015. "The Emergence of Lone Wolf Terrorism: Patterns of Behavior and Implications for Intervention." In *Terrorism and Counterterrorism Today*, vol. 20. Sociology of Crime, Law and Deviance. Emerald Group Publishing Limited. https://doi.org/10.1108/S1521-613620150000020005.
Spaaij, Ramon. 2010. "The Enigma of Lone Wolf Terrorism: An Assessment." *Studies in Conflict and Terrorism* 33 (9): 9.
Spaaij, Ramón, and Mark S. Hamm. 2015. "Key Issues and Research Agendas in Lone Wolf Terrorism." *Studies in Conflict & Terrorism* 38 (3): 167–78. https://doi.org/10.1080/1057610X.2014.986979.
Stake, Robert E. 1995. *The Art of Case Study Research*. Sage Publications.
Stampnitzky, Lisa. 2016. "The Emergence of Terrorism Studies as a Field." In *Routledge Handbook of Critical Terrorism Studies / Edited by Richard Jackson.* Routledge.
Stevenson, Megan. 2018. "Assessing Risk Assessment in Action." *Minnesota Law Review*, January 1. https://scholarship.law.umn.edu/mlr/58.
Strang, Sabrina, Christina Hoeber, Olaf Uhl, et al. 2017. "Impact of Nutrition on Social Decision Making." *Proceedings of the National Academy of Sciences* 114 (25): 6510–14. https://doi.org/10.1073/pnas.1620245114.
Strashnov, Amnon. 1994. *Justice under Fire*. Yediot Aharonot.
"The Occupation's Fig Leaf: Israel's Military Law Enforcement System as a Whitewash Mechanism | B'Tselem." n.d. Accessed March 8, 2024. http://www.btselem.org/publications/summaries/201605_occupations_fig_leaf.

Um, Eric van, and Daniela Pisoiu. 2015. “Dealing with Uncertainty: The Illusion of Knowledge in the Study of Counterterrorism Effectiveness.” *Critical Studies on Terrorism* 8 (2): 229–45. https://doi.org/10.1080/17539153.2014.981400.

UNCCT. 2021. “Countering Terrorism Online with Artificial Intelligence.” https://unicri.it/Publications/Countering-Terrorism-Online-with-Artificial-Intelligence-%20SouthAsia-South-EastAsia?utm_source=chatgpt.com.

Van Puyvelde, Damien, Stephen Coulthart, and M. Shahriar Hossain. 2017. “Beyond the Buzzword: Big Data and National Security Decision-Making.” *International Affairs* 93 (6): 1397–416. https://doi.org/10.1093/ia/iix184.

Vaswani, A. 2017. “Attention Is All You Need.” *Advances in Neural Information Processing Systems*. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Verbeek, Peter-Paul. 2006. “Materializing Morality: Design Ethics and Technological Mediation.” *Science, Technology, & Human Values* 31 (3): 361–80. https://doi.org/10.1177/0162243905285847.

Viterbo, Hedi. 2018. “Rights as a Divide-and-Rule Mechanism: Lessons from the Case of Palestinians in Israeli Custody.” *Law & Social Inquiry* 43 (3): 764–95. https://doi.org/10.1111/lsi.12270.

W Flores, Anthony, Kristin Bechtel, and Christopher Lowenkamp. 2016. “False Positives, False Negatives, and False Analyses: A Rejoinder to ‘Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks.’” *Federal Probation* 80 (September).

Wall, Christopher. 2024. “The (Non) Deus-Ex Machina: A Realistic Assessment of Machine Learning for Countering Domestic Terrorism.” *Studies in Conflict & Terrorism* 47 (6): 599–621. https://doi.org/10.1080/1057610X.2021.1987656.

Wang, Pei. 2019. “On Defining Artificial Intelligence.” *Journal of Artificial General Intelligence* 10 (2): 1–37. https://doi.org/10.2478/jagi-2019-0002.

Weber, Max. 1917. *Methodology of Social Sciences*. Transaction Publishers.

Wei, Wei, Xiang Li, Kun Liu, and Siyu Tan. 2024. “A Brief Analysis of Palantir Gotham: A Collaborative and Interactive Big Data Visualization Analysis Software Based on Dynamic Ontology.” *2024 10th International Conference on Big Data and Information Analytics (BigDIA)*, October, 769–76. https://doi.org/10.1109/BigDIA63733.2024.10808897.

Weill, Sharon. 2014. *The Role of National Courts in Applying International Humanitarian Law*. Oxford University Press.

Whittaker, Joe. 2022. “Rethinking Online Radicalization.” *Perspectives on Terrorism* 16 (4).

Whittaker, Joe, and Herath Charmin. 2021. “Online Radicalisation: Moving beyond a Simple Dichotomy.” *Terrorism and Political Violence* 0: 1–22.

Y., Colonel. 2018. “(Heb.), The Journey Towards Clarifying the Perception and Implementation of Intelligence and Operational Superiority in the Digital Era,.” *Modi in Halacha VeMaaseh, Big Data and Intelligence* 2.

Yavne, Lior. 2007. *BACKYARD PROCEEDINGS*. Yes Din. https://www.yesh-din.org/en/backyard-proceedings/.

Yin, Robert K. 2018. *Case Study Research and Applications: Design and Methods*. Sixth edition. SAGE.

Young, Iris Marion. 2006. “RESPONSIBILITY AND GLOBAL JUSTICE: A SOCIAL CONNECTION MODEL.” *Social Philosophy and Policy* 23 (1): 102–30. https://doi.org/10.1017/S0265052506060043.

Y.S. 2021. *The Human Machine Team: How to Create Synergy between Human & Artificial Intelligence That Will Revolutionize Our World*.

Završnik, Aleš. 2021. “Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings.” *European Journal of Criminology* 18 (5): 623–42. https://doi.org/10.1177/1477370819876762.

Zulaika, Joseba. 2009. *Terrorism: The Self-Fulfilling Prophecy*. University of Chicago Press. https://press.uchicago.edu/ucp/books/book/chicago/T/bo8103579.html.

Zulaika, Joseba. 2016. “The Real and the Bluff - On the Ontology of Terrorism.” In *Routledge Handbook of Critical Terrorism Studies*, edited by Richard Jackson. Routledge.
Zulaika, Joseba, and William A. Douglass. 1996. *Terror and Taboo: The Follies, Fables, and Faces of Terrorism*. Routledge.

Israeli Military Court Indictments Mentioned:

Indictment 2730/16 from the FOI request

Indictment 3231/16 from the FOI request

Indictment 2964/16 from the FOI request

Indictment 2966/16 from the FOI request

Indictment 4959/16 from the FOI request

Indictment 1345/16 from the FOI request

Indictment 1216/16 from the FOI request

Indictment 1338/16 from the FOI request

Indictment 1542/16 from the FOI request

Indictment 1383/16 from the FOI request

Indictment 1188/17 from the FOI request

Indictment 1345/16 from the FOI request

Indictment 1463/16 from the FOI request

Indictment 1622/16 from the FOI request

Indictment 1623/16 from the FOI request

Indictment 1159/16 from the FOI request

Indictment 3053/16 from the FOI request

Indictment 3055/16 from the FOI request

Indictment 3055/16 from the FOI request

Indictment 3114/16 from the FOI request

Israeli Supreme Court Cases:

HCJ 637/23 Israel Defence Forces Vs Ramati (IL 2023)

HCJ 5100/94 IL Vs DC 2115 (IL 1999)

HCJ 4211/91 El Masri Vs Israel 624 (IL 1993)

Israeli Military Court Cases:

Arrest appeal 3044/15/ The Military Prosecution Vs Abu Salim (published on Nevo.co.il)

Arrest appeal 1150/16/ The Military Prosecution Vs Al Harub (published on Nevo.co.il)

Arrest appeal 1222/16/ The Military Prosecution Vs Hamed (published on Nevo.co.il)

Arrest appeal 2933/16/ The Military Prosecution Vs Al Abu Absa (published on Nevo.co.il)

Appendix A - Ethics committee approval

Ollscoil Chathair Bhaile Átha Cliath
Dublin City University

DCU

Nery Ramati
School of Law and Government

Prof. Maura Conway
School of Law and Government

14th March 2022

REC Reference: **DCUREC/2022/026**

Proposal Title: **The Use of Preventive Artificial Intelligence Tools in Counter-Terrorism: The Case of the Israeli Security Agency (ISA) AI Tool**

Applicant(s): **Nery Ramati and Prof. Maura Conway**

Dear Colleagues,

Thank you for your application to DCU Research Ethics Committee (REC). Further to expedited review, DCU REC are pleased to issue approval for this research proposal. This approval is conditional on the DCU Data Protection Unit (DPU) approving the project and any related documentation, such as a data protection impact assessment (DPIA). Research should not begin until this is in place.

DCU REC's consideration of all ethics applications are dependent upon the information supplied by the researcher. This information is expected to be truthful and accurate. Researchers are responsible for ensuring that their research is carried out in accordance with the information provided in their ethics application.

Materials used to recruit participants should note that ethical approval for this project has been obtained from the Dublin City University Research Ethics Committee. Should substantial modifications to the research protocol be required at a later stage, a further amendment submission should be made to the REC.

Yours sincerely,

Dr. Melrona Kirrane

Dr. Melrona Kirrane
Chairperson
DCU Research Ethics Committee

DCU
Research & Innovation

Taighde & Nuálaíocht Tacaíocht
Ollscoil Chathair Bhaile Átha Cliath,
Baile Átha Cliath, Éire

Research & Innovation Support
Dublin City University,
Dublin 9, Ireland

T +353 1 700 8000
F +353 1 700 8002
E research@dcu.ie
www.dcu.ie