



JURIDISKA FAKULTETEN
vid Lunds universitet

Simone Isaksson Krantz

Att straffa det oförutsebara
– *En rättsdogmatisk analys av*
personlig skuld vid autonoma
systemfel

LAGF03 Rättsvetenskaplig uppsats

Kandidatuppsats på juristprogrammet

15 högskolepoäng

Handledare: Sverker Jönsson

Termin: Ht 2025

Innehåll

SUMMARY	1
SAMMANFATTNING	3
FÖRKORTNINGAR.....	5
1 INLEDNING	6
1.1 Bakgrund	6
1.2 Syfte och frågeställningar	7
1.3 Avgränsningar.....	8
1.4 Metod och material	8
1.5 Forskningsläge.....	10
1.6 Disposition.....	10
2 DEN SJÄLVKÖRANDE BILEN.....	12
2.1 Artificiell intelligens och maskininlärning	12
SAE – skalan	13
Beslutprocessen och black box-problematiken	13
3 DEN STRAFFRÄTTSLIGA ANSVARSLÄRAN	15
3.1 Brottsbegreppets konstruktion	15
Kravet på subjektiv och objektiv täckning	15
Täckningsprincipen	16
Vållande till annans död	16
3.2 Fundamenten i straffrätten.....	17
Skuldprincipen och konformitetskravet.....	17
3.3 Oaktsamhetsbegreppets konstruktion	17
Bedömningen av omedveten oaktsamhet	18
4 GRÄNSDRAGNING I RÄTTSPRAXIS	20
4.1.1 NJA 2006 s.228 – Spädbarnets död	20
<i>Aktsamhet i riskfylld verksamhet</i>	20
4.1.2 NJA 2023 s.29 – Mordet vid busshållplatsen.....	20
<i>Avsaknad av alternativa gärningsmän</i>	20

4.1.3	Högsta domstolen: B 2354-22 – Skjutningen vid Årsta torg	21
	<i>Ej ställt bortom rimligt tvivel.....</i>	<i>21</i>
5	RÄTTFÄRDIGANDE OMSTÄNDIGHETER	23
5.1	Ansvarsfrihetens yttersta gräns: <i>Social adekvans</i>	23
6	ANALYS	25
6.1	Krocken mellan black box och insiktsförmåga.....	25
6.2	Aktsamhetsnormen i riskfylld verksamhet.....	26
6.3	Bevisvärderingens moment 22.....	27
6.4	Värdering av social adekvans som nödfallsutgång.....	29
7	SLUTSATSER OCH RÄTTSPOLITISK DISKUSSION	31
7.1	Rättssäkerhet kontra det systematiska vakuumet.....	31
7.2	Konflikten mellan skuldprincipen och strikt ansvar.....	31
7.3	Transparens som lösning: återställande av skuldprincipen	32
	KÄLLFÖRTECKNING.....	34

Summary

The following thesis analyzes criminal liability in accidents caused by self-driving vehicles. These autonomous vehicles are governed by advanced artificial algorithms that make autonomous decisions, which challenges the fundamental requirements and principles of criminal law. Among other things, the maxim "nulla poena sine culpa" (no punishment without guilt) suffers when the decision-making process at the moment of the act lacks transparency and insight. In light of this, the thesis specifically analyzes the extent to which engineers or system developers, through negligence, can be held liable for causing the death of another pursuant to Chapter 3, Section 7 of the Swedish Penal Code (Brottsbalken). Through this, the thesis illustrates how the lack of transparency in AI systems, hereinafter referred to as the "black box problem," complicates the assessment of criminal liability.

Using a critical legal dogmatic method, it is analyzed whether current criminal law is sufficiently designed to handle technical innovation where decision-making processes remain hidden within complex systems. The elements of a crime in criminal law are constructed on the basis that the perpetrator is a natural person with an individual consciousness to make decisions. The artificial intelligence in self-driving vehicles creates a theoretical artificial perpetrator who lacks human consciousness and whose decision chains remain concealed.

A central conclusion is that the black box problem prevents the determination of personal guilt. According to the principle of conformity, criminal liability requires that the individual had the ability and opportunity to comply with the law. Since the engineers behind the systems lack insight into the system's decision chains, the prerequisites for establishing negligence through personal insight and blameworthiness are lacking. Furthermore, the standard of proof "beyond a reasonable doubt" is analyzed to illustrate the obstacles faced by the prosecutor due to the lack of insight into the systems. The thesis highlights reasoning from the Supreme Court regarding the requirement that alternative courses of events must be excluded for criminal

liability, which the inherent lack of transparency in AI systems prevents. The result of this is a liability vacuum in Swedish criminal law where the legislation cannot punish any human actor for the systems' shortcomings.

Finally, social adequacy (*social adekvans*) is analyzed as a potential ground for exclusion of liability, in order to illustrate how criminal law should relate to AI innovation. Despite the social benefits of AI, the thesis dismisses the applicability of the doctrine of social adequacy. When opaque decision-making processes in self-driving cars replace and erase human control, the risk-taking in traffic cannot be seen within the framework of what is socially accepted or justifiable.

Lastly, potential solutions for the legislator are presented to address the potential liability vacuum in the legislation. The thesis particularly highlights the introduction of requirements for Explainable AI (XAI). XAI creates the conditions to transform hidden algorithms into clear decision chains, which provides better opportunities to demand criminal liability and prevents artificial technology from eroding the purpose of criminal law.

Sammanfattning

Följande uppsats analyserar det straffrättsliga ansvaret vid olyckor orsakade av självkörande fordon. Dessa självkörande fordon styrs av avancerade artificiella algoritmer som fattar autonoma beslut, vilket utmanar straffrättens grundläggande krav och principer. Bland annat blir maximet ”*nulla poena sine culpa*” (inget straff utan skuld) lidande när beslutsprocessen i gärningsögonblicket saknar transparens och insyn. I ljuset av detta analyserar uppsatsen särskilt i vilken utsträckning som ingenjörer eller systemutvecklare, genom oaktsamhet, kan hållas ansvariga för vållande till annans död enligt 3 kap.7§ BrB. Uppsatsen åskådliggör genom detta hur bristen på transparens i AI-systemen, nedan benämnd black box-problematiken, försvårar den straffrättsliga ansvarsbedömningen.

Genom en kritisk rättsdogmatisk metod analyseras huruvida gällande straffrätt är tillräckligt utformad för att hantera teknisk innovation där beslutsprocesserna förblir dolda av de komplexa systemen. Rekvisiten inom straffrätten är konstruerade mot bakgrund av att gärningsmannen är en fysisk människa med ett eget medvetande att fatta beslut. Genom den artificiella intelligensen hos självkörande fordon skapas en teoretisk artificiell gärningsman som saknar mänskligt medvetande och vars beslutskedjor förblir dolda.

En central slutsats är att black box-problematiken hindrar fastställandet av personlig skuld. Enligt konformitetsprincipen förutsätter straffansvar att individen haft förmåga och tillfälle att rätta sig efter lagen. Då ingenjörerna bakom systemen saknar insyn i systemets beslutskedjor, brister förutsättningarna för att konstatera oaktsamhet genom personlig insikt och klander. Vidare analyseras beviskravet ”*ställt utom varje rimligt tvivel*” i syfte att åskådliggöra det hinder som åklagaren ställs inför till följd av bristen på insyn i systemen. Uppsatsen belyser resonemang från Högsta domstolen gällande att alternativa händelseförlopp måste kunna uteslutas för straffrättsligt ansvar, vilket AI-systemens inneboende brist på transparens hindrar. Resultatet av detta blir ett ansvarsvakuum i svensk straffrätt där lagstiftningen inte kan straffa någon mänsklig aktör för systemens brister.

Avslutningsvis analyseras social adekvans som en tänkbar ansvarsfrihetsgrund, i syfte att åskådliggöra hur straffrätten bör förhålla sig till AI-innovation. Trots samhällsnyttan av AI, avfärdar uppsatsen social adekvanslärans tillämplighet. När ogenomskinliga beslutsprocesser i självkörande bilar ersätter och raderar mänsklig kontroll kan risktagandet i trafiken inte ses inom ramen för socialt accepterat eller försvarligt.

Slutligen presenteras tänkbara lösningar för lagstiftaren i syfte att hantera det potentiella ansvarsvakuumet i lagstiftningen. Uppsatsen belyser särskilt införandet av krav på transparent AI (Explainable AI, XAI). XAI skapar förutsättningar att omvandla dolda algoritmer till tydliga beslutskedjor, vilket skapar bättre förutsättningar att utkräva straffrättsligt ansvar och motverkar att artificiell teknik urholkar straffrättens syfte.

Förkortningar

AI	Artificiell Intelligens
AI ACT	EU:s förordning om artificiell intelligens
BrB	Brottsbalken
CDPC	European Committee on Crime Problems (Europarådets expertorgan)
DDT	Dynamic Driving Task (Dynamiska köruppgifter)
DNN	Deep Neural Networks (Djupa neurala nätverk)
HD	Högsta domstolen
ML	Machine Learning (Maskininlärning)
NJA	Nytt juridiskt arkiv (Högsta domstolens rättsfallsamling)
SAE	Society of Automotive Engineers
SOU	Statens offentliga utredningar
SvJT	Svensk Juristtidning
XAI	Explainable Artificial Intelligence (Förklarbar AI)

1 Inledning

1.1 Bakgrund

År 2018 inträffar den första dödsolyckan med ett självkörande fordon i Arizona, USA. Fordonets system misslyckas med att identifiera den passerande fotgängaren, vilket resulterar i att det automatiska inbromsningssystemet aldrig aktiveras.¹ Händelsen illustrerar ett paradigmskifte inom straffrättens område: Hur ska det straffrättsliga ansvaret tillskrivas när en icke mänsklig aktör, i form av ett artificiellt intelligent system, orsakar en dödlig olycka? I vilken utsträckning kan ingenjörer och systemutvecklare hållas ansvariga för dessa systemfel?²

Denna problematik utmanar straffrättens fundamentala principer, i synnerhet kravet på personlig skuld och maximet *Nulla poena sine culpa*.³ Vidare har kravet på fri vilja som förutsättning för moraliskt och straffrättsligt ansvar genomsyrat rättsfilosofiska resonemang från antikens Aristoteles till moderna rättsfilosofier som Mill och Kant. Den gemensamma nämnaren för dessa resonemang är synen på tillräknelighet och förmågan att välja sina handlingar.⁴

Europarådets expertorgan adresserar även denna konflikt genom att belysa oaksamhetsdilemmat.⁵ Dilemmat åskådliggör svårigheterna med att utkräva straffrättsligt ansvar från mänskliga aktörer i tidigare led då de självkörande fordonens beslutskedjor är dolda.⁶ Rådet adresserar även problematiken

¹D. Wakabayashi, 'Self-Driving Uber Car Kills Pedestrian in Arizona', *The New York Times*, 20 mars 2018.

² Utrecht University, Sven Nyholm, 'Unboxing the black box of AI', 2021.

³ SOU 2012:17 s.522. "Inget straff utan skuld".

⁴ R. Johnson & A. Cureton, 'Kant's Moral Philosophy', *The Stanford Encyclopedia of Philosophy*, 2022; C. Macleod, 'John Stuart Mill', *The Stanford Encyclopedia of Philosophy*, 2020; T. O'Connor & C. Franklin, 'Free Will', *The Stanford Encyclopedia of Philosophy*, 2021.

⁵ CDPC, Discussion Paper on Criminal Liability Related to AI systems, CDPC(2024)09, 2024, s. 7, p. 2b.

⁶ *ibid*, O. Picken (red.), 'Explainable AI: Addressing Challenges for Machine Learning Implementation', *Oxford Global*, 2022.

kring att AI i dagsläget inte kan betraktas som en moralisk människa och kan därigenom inte vara mottagare av straff eller placeras i fängelse.⁷

Mot denna bakgrund framträder en tydlig konflikt: Ett självkörande fordon kan objektivt sett uppfylla rekvisiten för en brottslig gärning, *actus reus*, men saknar genom sin tekniska natur, möjligheten att uppfylla det subjektiva rekvisitet på skyldigt sinne, *mens rea*.⁸ Den snabba utvecklingen av artificiell intelligens och självkörande fordon ställer straffrätten inför ett vägval; Vem kan med stöd av nuvarande lagstiftning hållas straffrättsligt ansvarig för dödliga olyckor orsakade av en artificiell algoritm utan mänsklig vilja?

1.2 Syfte och frågeställningar

Uppsatsen författas i syfte att utreda skuldprincipens tillämplighet vid dödliga olyckor orsakade av självkörande fordon på nivå 5.⁹ Fokus ligger på att belysa hur AI-systemens bristande transparens förhåller sig till rekvisitet omedveten oaktsamhet och möjligheten att utkräva straffansvar av individer bakom tekniken. Analysen syftar vidare till att fastställa om gällande rätt är tillräcklig eller om ett rättsligt vakuum uppstår, samt att presentera potentiella lösningar ur ett rättspolitiskt perspektiv. Följande frågeställningar besvaras för att uppnå dessa syften:

- *Hur påverkar AI-systemens brist på transparens och förutsägbarhet bedömningen av personlig skuld, särskilt avseende kraven på insikt vid omedveten oaktsamhet?*
- *I vilken utsträckning uppstår ett straffrättsligt ansvarsvakuum vid tillämpningen av gällande rätt på nivå 5-fordon, och kan läran om social adekvans fungera som en rättfärdigande grund för systemfel?*
- *Vilka konsekvenser för rättssäkerheten medför ett införande av strikt ansvar respektive krav på teknisk transparens (XAI) i förhållande till skuldprincipen och beviskravet i brottmål?*

⁷ European Committee on Crime Problems (CDPC), *Discussion Paper on Criminal Liability Related to AI systems*, CDPC(2024)09, 2024, s.6 p.1.

⁸ Maastricht Institute for Criminal Sciences (MICS), *Criminal behavior and accountability of artificial intelligence systems*, Maastricht University, 2023, s.14 f.

⁹ (Denna nivå avser fordon med full autonomi, där ingen mänsklig förare behövs för att framföra fordonet).

1.3 Avgränsningar

Vissa avgränsningar har gjorts i syfte att möjliggöra en fördjupad analys inom de givna ramarna. Uppsatsens tekniska fokus ligger på att behandla ansvarsfrågan vid olyckor orsakade av fordon på nivå 5 i SAE-skalan, där det råder full autonomi. Fordon på tidigare nivåer lämnas därför utanför uppsatsen. Vidare avgränsas undersökningen till det mänskliga ansvaret i tidigare led, med primärt fokus på systemutvecklare och ingenjörer.

Uppsatsens juridiska fokus ligger på det personliga skuldansvaret. Centralt för analysen är den subjektiva sidan av brott och skuldformen culpa, där ”omedveten oaktsamhet” hamnar i fokus. Den objektiva sidan av brott behandlas inte inom ramen för uppsatsen då analysen utgår från premissen att de objektiva förutsättningarna är uppfyllda i teorin. Uppsatsen har 3 kap.7§ BrB. ”vållande till annans död”: som juridisk utgångspunkt för att belysa och analysera det straffrättsliga skuldansvaret. Analysen är vidare utformad genom att rikta in sig på orsaksledet i oaktsamhetsbedömningen i syfte att åskådliggöra huruvida individerna tidigare i ledet hade förmåga och tillfälle att inse risken.

1.4 Metod och material

För att uppnå syftet med uppsatsen valde jag den rättsdogmatiska metoden. Denna metod karaktäriseras av en konkret formulerad problemframställning¹⁰ och kan beskrivas som en vetenskaplig rekonstruktion av gällande rättssystem.¹¹

Metoden innefattar även tillämpning av allmänt accepterade rättskällor¹² såsom lagstiftning, lagförarbeten, och rättspraxis vilka inom svensk rätt tillmäts formell auktoritet inom rättskälleläran.¹³ Inom rättsdogmatikens ramar

¹⁰ Kleineman (2025) s.28.

¹¹ Kleineman (2025) s.30.

¹² Kleineman (2025) s. 25. Se även Jareborg (2004) s. 4.

¹³ Jfr. Jareborg (2004) s. 9 & Kleineman (2025) s. 34.

möjliggörs även en kritisk analys innebärande att studera och analysera brister hos gällande rätt ur en kritisk synvinkel.¹⁴ Utöver uppsatsens deskriptiva de lege lata karaktär, valde jag att föra en kritisk rättsdogmatisk de lege ferenda diskussion av bristerna i nuvarande straffrättslig lagstiftning, i ljuset av AI och självkörande fordon.

Då specifik straffrättslig lagstiftning för autonoma bilar på nivå 5 saknas, används en analogisk lagtillämpning av generella straffrättsliga principer från HD:s praxis. Metoden innebär att jag identifierar och analyserar avgöranden från HD som uppvisar tydliga likheter med min uppsats problemframställning, som exempelvis personlig insiktsförmåga vid komplexa händelseförlopp samt vilka beviskrav som ställs vid indiciekedjor.¹⁵ Genom detta avser jag att pröva huruvida mina frågeställningar kan besvaras med hänsyn till nuvarande lagstiftning, (*de lege lata*). Rättsreglerna tillämpas dock restriktivt i syfte att bibehålla rättssäkerheten, och belyser därav enbart hur principerna *skulle kunna* tillämpas.¹⁶

Uppsatsen tar sin utgångspunkt i betänkandet SOU 2016:28 som visserligen fokuserar på försöksverksamheten av självkörande fordon. Dock anser jag att betänkandet belyser intressanta aspekter och resonemang som blir högst aktuella i samband med framtida lansering av självkörande fordon.

Utöver nationell rätt, integrerar jag även EU-rättsliga perspektiv genom AI-förordningen (AI Act), som belyser kraven på transparens och riskhantering som ställs på mänskliga aktörer. Det svenska materialet kompletteras vidare med EU-rättsliga etiska riktlinjer i syfte att åskådliggöra den framväxande aktsamhetsnormen inom artificiell intelligens. Vidare implementeras material från teknologisk facklitteratur och internationell forskning kring främst black box-problematiken, djupa neurala nätverk och maskininlärning, i syfte att skapa en grundläggande förståelse för de artificiella systemens uppbyggnad. Vidare implementerar jag standarder från Society of Automotive Engineers i syfte att klargöra vilken nivå av autonomi hos fordonet som

¹⁴ Kleinman (2025) s. 49.

¹⁵ Kleinman (2025) s. 26.

¹⁶ Kleinman (2025) s. 37 f. och s. 44.

min uppsats tar sin utgångspunkt i samt belysa den tekniska definitionen av termen autonomi. Slutligen redogör jag för ett rättspolitiskt perspektiv, *de lege ferenda*, med primärt syfte att analysera gällande rätt mot framtida åtgärder för lagstiftning.¹⁷

1.5 Forskningsläge

Forskningsläget präglas av spänningen mellan AI-utveckling och traditionell ansvarslära. SOU 2018:16 belyser hur autonoma körsystem utmanar konformitets och skuldprincipen, särskilt när kontrollen överförs från människa till maskin. Utredningen aktualiserar frågan om huruvida tillverkare, passage- eller importörer besitter den faktiska förmågan att påverka händelseförloppet.¹⁸

Sveriges Advokatsamfund betonar vikten av legalitet och förutsebarhet, där beviskravet ”ställt utom varje rimligt tvivel” förutsätter en tydlig definition av ”förare” begreppet. Detta i syfte att undvika oförutsedda sanktioner mot enskilda.¹⁹ Inom doktrinen betonar Ulf Lundqvist att AI-systemens brist på fri vilja försvårar tillämpningen av skuldkravet, och ser därför strikt ansvar för ägare eller tillverkare som en lösning på denna problematik.²⁰

Slutligen adresserar EU:s AI Act och EU-kommissionens expertgrupp indirekt black box-problematiken genom krav på teknisk transparens och förklarbarhet.²¹ Enligt dessa riktlinjer är insyn och mänsklig kontroll nödvändiga förutsättningar för att kunna utkräva straffrättsligt ansvar.²²

1.6 Disposition

¹⁷ Kleineman (2025) s.44 f.

¹⁸ SOU 2018:16 s. 554 f.

¹⁹ Sveriges Advokatsamfund, *Remissvar avseende SOU 2018:16*, R-2018/0777, 2018, s. 1–3.

²⁰ Ulf Lundqvist, 'Artificiell Intelligens - rättsordning och rättstillämpning', *SvJT*, 2020, s. 388.

²¹ Förordning (EU) 2024/1689 (AI-förordningen), art. 13.1–14.1.

²² EU-kommissionen, *Ethics guidelines for trustworthy AI*, High-Level Expert Group on Artificial Intelligence, 2019, s.14 p.53 och s.16 p.4.

Uppsatsen är indelad i sju olika kapitel. Nästkommande kapitel redogör för det tekniska ramverket i självkörande fordon, där särskilt fokus riktas på att belysa maskininlärning och black box-problematiken. I kapitel tre behandlas den straffrättsliga ansvarsläran med fokus avseende skuldprincipen, personligt skuldansvar och oaktsamhetsbegreppets konstruktion. Kapitel fyra belyser i sin tur relevant praxis från Högsta domstolen i syfte att exemplifiera HD:s bedömningsgrunder inom straffrätten. Kapitel fem preciserar innebörden av läran om social adekvans som en potentiell rättfärdigande omständighet vid fel i systemet. Dessa delar sammanförs i kapitel sex, vilket utgör uppsatsens analys. Här prövas den tekniska autonomin mot straffrättens krav på personlig skuld och insiktsförmåga, samtidigt som hinder för bevisningen och social adekvans analyseras. Uppsatsen avslutas i kapitel sju med en rättspolitisk diskussion kring framtida lösningar samt mina avslutande slutsatser.

2 Den självkörande bilen

2.1 Artificiell intelligens och maskininlärning

År 1955 myntar John McCarthy, forskare vid Stanford University, termen Artificiell Intelligens, med följande definition; ”*Den delen av vetenskapen och tekniken som skapar intelligenta maskiner.*”²³

Artificiell intelligens i kontexten självkörande bilar syftar till den delen av ingenjörskonst som ligger till grund för skapandet av autonoma system.²⁴ Dessa autonoma system är utformade i syfte att efterlikna mänskliga egenskaper, exempelvis förmågan att fatta självständiga beslut i relation till sin omgivning.²⁵ Termen autonomi i denna kontext innebär således en helt egen typ av autonomi. Detta beror på att autonoma system besitter förmågan att besluta om sekvenser av steg för att uppnå ett mål, helt genom självständig planering utan förprogrammerad detaljstyrning. Detta möjliggörs genom maskininlärning (ML), vilket syftar till den delen av AI som studerar hur systemen kan förbättra sina mönster för inlärning. Detta sker genom att erfarenhet och data läggs till grund för att bedöma hur perception, kunskap och handlingsförmåga ska förbättras. Avsaknad av maskininlärning i AI-systemet medför att systemet bibehåller sin karaktär som statisk. Detta skiljer exempelvis en fabriksrobot från en självkörande bil. Roboten i fabriken å ena sidan, agerar statiskt i en oföränderlig miljö, medan den självkörande bilen å andra sidan, självständigt anpassar sig efter sin omgivning i en föränderlig miljö.²⁶

Framväxten av den nya tekniken och AI-system med dess maskininlärning som liknar den mänskliga hjärnan i allt högre grad,²⁷ nödvändiggör att högre

²³ Stanford University (HAI), 'Brief Definitions of Key Terms in AI', *Stanford Institute for Human-Centered AI*, 2022.

²⁴ Imagination Technologies, "Artificial Intelligence in Autonomous Vehicles".

²⁵ ibid Stanford University (HAI).

²⁶ ibid Stanford University (HAI).

²⁷ Brown University, "Researchers uncover similarities between human and AI learning", *Brown University News*, 4 september 2025.

krav ställs på ansvarsfull design och noggrann planering vid såväl implementering som systemutveckling.²⁸

SAE – skalan

För att fastställa vilken grad av autonomi som ett självkörande fordon besitter, tillämpas den internationellt erkända SAE-skalan.²⁹ Denna skala anger rådande branschstandard för att klassificera och definiera graden av autonomi i ett motorfordon. Skalan indelas i 6 nivåer, rangordnade från 0 – 5, där nivå 0 avser helt manuella fordon där en mänsklig förare hela tiden krävs för interaktion och övervakning i körningen. Nivå 5 avser i stället den högsta nivån av autonomi, där full självständighet råder genom att bilen hanterar alla delar av körningen utan någon förare närvarande i bilen.³⁰ Centralt för SAE-skalans indelning är således fordonets förmåga att utföra självständiga dynamiska köruppgifter. (DDT) i förhållande till mänsklig involvering i körningen.³¹

Beslutprocessen och black box-problematiken

I det självkörande fordonet finns avancerade AI-system som fyller funktionen som den självkörande bilens hjärna. Det är här som logiken kring beslut byggs upp genom maskininlärning, vilket sker inuti djupa neurala nätverken (DNN). I dessa nätverk lär sig bilen att bygga upp sin förståelse genom flera lager av matematiska omvandlingar som inte kan översättas till mänskliga ord eller regler.³² Bilens AI-system lär sig genom dessa neurala nätverk att identifiera komplexa mönster i stället för att dessa mönster och regler programmeras in på förhand. Detta resulterar i sin tur i att bilens agerande i trafiksituationer styrs av bilens egen interna logik som systemet själv

²⁸ IEEE Standards Association, ”The IEEE Global Initiative 2.0 on Ethics of Autonomous and Intelligent Systems”, 2025.

²⁹ SAE International, *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems J3016_202104*, 2021, SOU 2016:28 s.37 f och 38 f.

³⁰ S. Agarwal & J. De Freitas, ”Acceptance of Automated Vehicles Is Lower for Self than Others”, *Harvard Business School Working Knowledge*, 2024.

³¹ SAE International (2021); *Driving the Future – Levels of Autonomous Vehicles Explained*, 2021, University of Oxford, *Key points of investigation and analysis on traffic accidents involving intelligent vehicles*, s. 3.

³² University of Oxford, Marta (2017), Jfr Knight (2017), Chalmers, Lundqvist & de Richelieu, (2019) s. ix.

har skapat och byggt upp. Bilen skapar således en egen intuitiv förståelse för bilkörning.³³ Denna typ av artificiellt skapad inlärning ställer höga krav på transparens, trovärdighet och ansvarsskyldighet.³⁴

Det är denna självständiga logik och maskininlärning som banar väg för den ökända **black box-problematiken**. Att bilen skapar sin egen logik inuti dessa djupa neurala nätverken (DNN) medför att beslutsprocessen blir ogenomskinlig, vilket skapar tekniska svårigheter när man i efterhand vill försöka rekonstruera och finna en förklaring till varför bilen tog ett specifikt beslut i en trafiksituation.³⁵ Denna avsaknad av transparens i systemens algoritmer utgör black box-problematiken, där inte ens ingenjörerna bakom AI-systemen kan se vad som har blivit fel eller hur den internt skapade logiken resonerade i sina beslut.³⁶

³³ Knight (2017), F, Utesch m.fl., 'Towards behaviour-based testing to understand the black box of autonomous cars', 2020, s. 1-2 f.

³⁴ University of Oxford, Marta (2017).

³⁵ Utesch m.fl. (2020), s. 1-2 f, University of Oxford, 'Key points of investigation and analysis on traffic accidents involving intelligent vehicles', *Oxford University Research Archive*, 2021.

³⁶ Knight (2017), Aine Josephine Tyrell och Ben Douglas-Jones, *AI in Criminal Law*, Bloomsbury Professional Law Insight, 2025, s. 94.

3 Den straffrättsliga ansvarsläran

3.1 Brottsbegreppets konstruktion

Inom svensk straffrätt definieras brott som en i lag beskriven rättsstridig och klandervärd handling för vilken straff är föreskrivet.³⁷ Detta utgör ett uttryck för legalitetsprincipen, vilket innebär inget brott utan lag.³⁸ Bedömningen av om en handling uppfyller denna beskrivning avgörs med utgångspunkt i en tvåstegsbedömning; en handling är inte ett brott om inte även kravet på skuld är uppfyllt.³⁹ Centralt för denna uppsats är brott begångna genom omedveten oaktsamhet. Av 1. kap 2§ BrB. framhålls att ”*en gärning, om inte annat är särskilt föreskrivet, anses som brott endast då den begås med uppsåt.*” Ansvar för oaktsamhetsbrott fordrar således att brottsbeskrivningen uttryckligen anger att brott genom oaktsamhet är straffbara.⁴⁰

Kravet på subjektiv och objektiv täckning

Den svenska straffrättens syn på brott och brottsliga handlingar ses som en konstruktion av två delar.⁴¹ Å ena sidan den objektiva delen av brottet, innefattande brottsbeskrivningen av den konkreta gärningen och omständigheterna i övrigt. Å andra sidan, med stöd av täckningsprincipen, den subjektiva sidan som innefattar gärningsmannens subjektiva inställning, dvs kognitiva insikt, vilja och mål i relation till de objektiva omständigheterna.⁴² Den objektiva sidan (*actus reus*) motsvarar brottsbeskrivningen, dess följd och ett krav på orsakssamband (kausalitet).⁴³ Den subjektiva sidan (*mens rea*) kräver i sin tur att gärningsmannen har handlat med uppsåt eller oaktsamhet (*culpa*) i förhållande till de objektiva rekvisiten.⁴⁴

³⁷ J. Tapani & M. Tolvanen, (2016) s. 9, 1 kap. 1 § BrB.

³⁸ Prop. 1993/94:130 s. 17 f. och SOU 2018:16 s. 554.

³⁹ 1 kap. 2 § BrB.

⁴⁰ P. Asp, kommentar till 1 kap. 2 § BrB.

⁴¹ SOU 2018:16 s. 555 och prop. 1994/95:23 s. 54.

⁴² Jfr 1 kap. 1 § BrB, SOU 2018:16 s. 553–554, prop. 1993/94:130 s. 21 f.

⁴³ U. Lundqvist, 'Om betydelsen av komparativa studier för rättstillämpningen och framväxten av en global rättsordning', *SvJT*, 2019, s. 818, Asp m.fl. (2013) s. 58; SOU 2018:16 s. 555–557.

⁴⁴ Ibid.

Täckningsprincipen

Centralt för denna uppsats är den så kallade *täckningsprincipen*, vilken är en bärande princip inom straffrätten som anger att gärningsmannens oaktsamhet eller uppsåt måste täcka samtliga objektiva rekvisit i brottsbeskrivningen.⁴⁵ HD åskådliggör i NJA 2016 s.702. hur bedömningen av denna princip genomförs med följande uttalande; ”När en domstol har konstaterat att de objektiva förutsättningarna för en gärning är uppfyllda bör den första ställning till vilket krav på subjektiv täckning som gäller i det enskilda fallet.”⁴⁶

Vållande till annans död

Uppsatsen fäster stor vikt vid brottet vållande till annans död som återfinns i 3 kap 7§ BrB, vilket kriminaliserar den individ som genom oaktsamhet vållar en annan människas död. Denna straffrättsliga bestämmelse skildrar såväl den objektiva gärningsbeskrivningen som de subjektiva rekvisit som fordras för straffansvar.⁴⁷ En väsentlig skillnad vid bedömningen av brottet vållande till annans död, i kontrast till uppsåtliga brott, är dock att fokus inte ligger på gärningsmannens *faktiska vilja för handlingen*. Ansvar vilar i stället på bedömningen av dels vad gärningsmannen *borde ha insett* och dels bedömningen av huruvida handlingen innebar ett otillåtet risktagande ur samhällelig synpunkt.⁴⁸ HD stadgar i NJA 2020 s. 397 att en gärning som avviker från ett aktsamt agerande, utgör straffbar oaktsamhet.⁴⁹

Det bör dock framhållas att lagtexten i 3 kap.7§ BrB. endast stadgar att oaktsamhet är straffbar, men anger inte någon närmre definition av termen oaktsamhet. Det blir därför relevant att söka ledning i straffrättens allmänna principer och i doktrinen.⁵⁰

⁴⁵ Asp m.fl. (2013) s. 64 och 324.

⁴⁶ NJA 2016 s.702, p.18.

⁴⁷ Suzanne Wennberg, *Introduktion till straffrätten*, 13 uppl., Norstedts Juridik, Stockholm, 2020, s. 29–30.

⁴⁸ Lexino, kommentar till 3 kap.7§ BrB, avsnitt 2.2.

⁴⁹ Se NJA 2020 s. 397; Domstolen uttalar i p.18: ”i regel bör man utgå ifrån att andra personer beter sig någorlunda rationellt och t.ex. följer anvisningar och tar ansvar för sin egen hälsa och säkerhet”. (Den s.k. förtroendegrundsatsen).

⁵⁰ Se 1 kap. 1–2 §§ BrB.

3.2 Fundamenten i straffrätten

Skuldprincipen och konformitetskravet

Den svenska straffrätten vilar på grundmaximen; *Nulla poena sine culpa*, vilket är det latinska uttrycket för: ”*inget straff utan skuld*.”⁵¹ Principen har sitt ursprung i synen på människan som en självständig varelse med fri vilja och förmåga att fatta rationella val.⁵² En förutsättning för att hålla en individ straffrättsligt ansvarig, enligt skuldprincipen, är att individen kan klandras personligen för sina handlingar.⁵³ Vid oaktsamhetsbrott innebär detta att gärningsmannen, trots tillfälle och insikt, väljer att avvika från rådande aktsamhetsnorm.⁵⁴

Nära sammankopplat med skuldprincipen är **konformitetsprincipen**.⁵⁵ Principen fyller funktionen som en grundsats inom den moderna straffrättsliga ansvarsläran och innebär att en individ kan hållas straffrättsligt ansvarig om han hade *förmåga och tillfälle* att rätta sig efter lagen, men av någon anledning inte gjorde det.⁵⁶ Skuldformen oaktsamhet är i denna kontext ett nödvändigt rekvisit att analysera för att bedöma huruvida den subjektiva sidan av brottet är uppfyllt samt förutsättningarna för personligt ansvar.⁵⁷ Dessa två principer blir särskilt relevanta i uppsatsens analys där principernas funktion och syfte ställs på sin yttersta spets vid användningen av självkörande bilar med oförutsägbara ageranden.⁵⁸

3.3 Oaktsamhetsbegreppets konstruktion

⁵¹ NJA 2011 s. 563 p. 20.

⁵² Nils Jareborg, *Straffrättsideologi*, Iustus, Uppsala, 1992, kap.10; Asp m.fl. (2013), s. 269.

⁵³ NJA 2011 s. 563 p. 20.

⁵⁴ Asp m.fl. (2013) s. 270.

⁵⁵ SOU 2002:3 s. 230 f.

⁵⁶ NJA 2012 s. 564 p. 18, Asp m.fl. (2013) s. 270, SOU 2018:16 s. 553.

⁵⁷ Asp m.fl. (2013) s. 58–59.

⁵⁸ Jfr SOU 2018:16 s. 554 yttrande: ”Ett automatiskt körsystem utmanar konformitetsprincipen och skuldprincipen och ifrågasätter vem som kan ha skulden och vara oaktsam.”

Vid trafikolyckor orsakade av autonoma system, blir sällan uppsåt relevant varför ansvarsprövningen istället riktar in sig på **oaktsamhet (culpa)**. I doktrinen definieras termen oaktsamhet som en klandervärd brist i gärningsmannens vilja eller förmåga att rätta sig efter och klanderfritt följa gällande oaktsamhetsnormer.⁵⁹ Straffrätten delar in culpabegreppet i två kategorier;⁶⁰

Å ena sidan den medvetna oaktsamheten, *Culpa Luxuria*, som föreligger då gärningsmannen inser risken för en viss följd eller omständighet, men trots denna insikt, litar på att följden inte inträffar. Exempelvis köra högt över hastighetsbegränsningen trots vetskapen om riskerna. Gärningsmannen hyser med andra ord en förhoppning om att risken för skada inte förverkligas.⁶¹

Å andra sidan den omedvetna oaktsamheten, *Culpa negligentia*, som i stället föreligger när gärningsmannen inte har insett risken för en viss omständighet eller följd, men där han borde ha insett detta. Gärningsmannen borde med andra ord ha tänkt efter eller vidtagit kontrollåtgärder för att inse risken givet de aktuella omständigheterna och rättat sitt handlande därefter.⁶² Den klandervärda bristen ligger i att gärningsmannen borde ha insett risken. Det förutsätts dock inte att han eller hon borde ha insett att en omständighet **verkligen** förelåg, utan endast att det fanns en **risk**.⁶³

Bedömningen av omedveten oaktsamhet

Domstolen har i ett antal rättsfall åskådliggjort vilka grunder som de stödjer sin bedömning på vid omedveten oaktsamhet. En tvåstegsanalys genomförs i syfte att upprätthålla konformitetsprincipen och därigenom bringa klarhet i

⁵⁹ Wennberg, S., *Introduktion till straffrätten*, 13 uppl., 2020, s. 43–44; Wennberg, S. m.fl., *Brottsbalken: En kommentar*, till BrB 3:7 s.2 och 3.

⁶⁰ NJA 2022 s 826 p. 15.

⁶¹ *ibid.* p. 16.

⁶² *ibid.* p. 17; Asp m.fl. (2013) s. 315.

⁶³ Tapani & Tolvanen (2016) s. 136 och 138 f, SOU 2018:16 s. 557.

huruvida handlingen är att betrakta som klandervärd eller inte. Omedveten oaktsamhet prövas i två led;⁶⁴

Orsaksledet

Det första ledet i prövningen avser att undersöka vad gärningsmannen kunde ha gjort för att nå relevant insikt eller misstanke om risken.⁶⁵ Domstolens prövning lägger här stor vikt vid faktiska åtgärder som krävs för att komma till insikt kring den önskvärda aktsamheten och som gärningsmannen borde ha iakttagit.⁶⁶ Vanligtvis rör det sig om enkla handlingar i syfte att uppnå denna insikt såsom exempelvis att utöva självkontroll, iaktta information, tänka efter och visa tillräcklig uppmärksamhet i den givna situationen.⁶⁷

Klandervärdhetsledet

Det andra ledet avser en bedömning om det underlåtna handlandet, som identifierats i det första ledet, också är rättsligt klandervärdt. Denna bedömning sker genom att pröva huruvida det kan krävas att gärningsmannen **borde** ha utnyttjat sin förmåga och sitt tillfälle att nå insikt om risken.⁶⁸ I korthet utgör detta en normativ bedömning kring om gärningsmannen kan straffas för att han inte har gjort vad som rimligen har kunnat krävas för att följa aktsamhetsnormen.⁶⁹

⁶⁴ Asp m.fl. (2013) s. 315; Jfr NJA 2022 s.826 p. 5 och HovR domskäl: "*Oaktsamheten kan vara antingen medveten eller omedveten. Någon kvalificerad form av oaktsamhet krävs alltså inte. En allmän oaktsamhetsbedömning ska göras.*"

⁶⁵ Asp m.fl. (2013) s. 315 f.

⁶⁶ NJA 2022 s. 826 p. 16 och 17. (Domstolen behandlar här skillnaden mellan medveten och omedveten oaktsamhet.)

⁶⁷ NJA 2022 s. 826 p. 18-20, se även "Laserpekaren" NJA 2011 s. 725 p. 6 och 7; Jfr även NJA 2020 s. 397, (Tingsrättens domskäl gällande termen *gärningssculpa*).

⁶⁸ NJA 2011 s. 725 p.8 (Gällande bedömningen av oaktsamhet.), Asp m.fl. (2013) s. 315 f och s.317 f.

⁶⁹ *ibid.*; Asp och Ulväng, (2019) s. 50; Jfr prop. 2017/18:177 s. 88.

4 Gränsdragning i rättspraxis

Högsta domstolen har genom en serie avgöranden preciserat de strikta beviskrav som uppställs vid indiciebevisning samt belyst gränsen för straffbar oaktsamhet. Nedanstående prejudikat har en central betydelse för uppsatsens analys, eftersom dessa rättsfall genom analogier möjliggör en förståelse för hur det straffrättsliga ansvaret kan tillskrivas mänskliga aktörer vid händelseförlopp orsakade av självständiga AI-system

4.1.1 NJA 2006 s.228 – Spädbarnets död

Aktsamhet i riskfylld verksamhet

Omständigheterna i målet rör åtalet mot en sjuksköterska för vållande till ett spädbarns död, då hon av misstag ger en infusionslösning med tio gånger högre styrka än ordinerat. HD resonerar kring huruvida sjuksköterskan kan hållas straffrättsligt ansvarig för oaktsamhet då hon av misstag har vållat barnets död. HD anger att vid fastställande av ansvar, krävs att oaktsamhet föreligger. Detta innefattar en samlad bedömning av huruvida omständigheten i fallet är så pass klandervärd att det medför straffbar oaktsamhet. HD hävdar vidare att den omständigheten att droppet skulle sättas in på ett spädbarn borde ha föranlett ”*särskild uppmärksamhet och noggrannhet*” hos sjuksköterskan. HD betonar särskilt att verksamhetens art, det vill säga en riskfylld miljö, medför en aktsamhet som generellt sett är högre än i andra situationer. Ett sjukhus där läkemedelshantering medför särskilda risker och höga krav på hanteringen anses falla inom detta. HD fastställde hovrättens domslut och ansåg att sjuksköterskans handlande var så pass klandervärd att det innefattade straffbar oaktsamhet och hon dömdes därigenom för vållande till annans död. Rättsfallet åskådliggör vilken aktsamhet som krävs inom riskfylld verksamhet.⁷⁰

4.1.2 NJA 2023 s.29 – Mordet vid busshållplatsen.

Avsaknad av alternativa gärningsmän

⁷⁰ NJA 2006 s.228.

Målet rör händelseförloppet där en individ har avfyrat ett antal skott mot en person som sedermera avlider till följd av sina skador. I rättegången presenterar åklagaren en rad fristående bevis med samverkande effekt i syfte att styrka åtalet mot den tilltalade. Bevisen utgjordes av en indiciekedja där DNA-bevisning säkrades på patronhylsor från brottsplatsen samt på klädesplagg. HD ansåg mot bakgrund av bevisningens sammanvägda styrka, att det gick att ställa bortom rimligt tvivel, att den tilltalade bedömdes vara gärningsman till mordet.⁷¹ HD betonade särskilt att det genom den presenterade bevisningen, *var omöjligt* att komma fram till alternativa gärningsmän vilket styrkte åtalet mot den tilltalade.⁷² Rättsfallet åskådliggör vikten av att domstolen kan utesluta alternativa händelseförlopp för att uppfylla beviskravet.

4.1.3 Högsta domstolen: B 2354-22 – Skjutningen vid Årsta torg

Ej ställt bortom rimligt tvivel

Rättsfallet rörde en skottlossning riktad mot två ordningsvakter som befann sig i en bil. Riksåklagaren rubricerade händelsen som mordförsök och menade att de tilltalade hade införskaffat skarpladdade vapen i syfte att beröva ordningsmännen livet. Av målet framhölls följande beviskrav; när det finns två personer som i samförstånd påstås ha utfört handlingen, fordras bevisning på individnivå, där dessa medverkat till brottet på ett sätt som betraktas som medgärningsmän. I kontrast till HD:s bedömning i mordet vid busshållplatsen, ogillar HD åtalet i detta mål. Enligt HD var den framförda bevisningen inte tillräckligt stark för att tydligt fastställa vilken roll i skottlossningen som de båda tilltalade hade samt om det fanns andra delaktiga personer. Enligt HD måste bevisningen vara av så individualiserad karaktär att det kan ställas bortom rimligt tvivel att de tilltalade är skyldiga. Det räcker inte med att de tilltalade sannolikt är skyldiga. Åklagaren misslyckades i fallet att styrka bortom

⁷¹ Mål B 5952-21, s. 20 p. 77.

⁷² Mål B 5952-21, s. 19 p. 69.

varje rimligt tvivel och de båda tilltalade friades.⁷³ Rättsfallet illustrerar kravet på individualiserat ansvar, vilket blir relevant för kommande analys.⁷⁴

⁷³ Mål B 2354-22 s. 19 p. 56 och 57.

⁷⁴ Sveriges Domstolar, *'Två avgöranden om bevisvärdering i brottmål'*, 2023.

5 Rättfärdigande omständigheter

5.1 Ansvarsfrihetens yttersta gräns:

Social adekvans

Högsta domstolen betonar att det vid sidan av brottsbalkens tydliga undantagsregler i 24 kap BrB, även finns möjligheten att bedöma en gärning som tillåten med stöd av oskrivna regler som läran om social adekvans.⁷⁵ Detta då det finns ett samhälleligt intresse av att gärningen ska vara tillåten.⁷⁶ I praktiken innebär detta att vissa annars straffbelagda gärningar *undantas* från det kriminaliserade området, trots att gärningarna uppfyller rekvisiten för att ses som brottsliga.⁷⁷ Detta möjliggörs genom läran om social adekvans, det vill säga den *oskrivna* ansvarsfrihetsgrunden och rättfärdigas av allmänna överväganden av vad som ur samhällelig synpunkt betraktas som socialt önskvärt.⁷⁸

Ett konkret exempel där läran om social adekvans får betydelse är inom sjukvård och idrottsutövning. Läkare som genomför operationer dagligen inom ramen för sin sjukvårdsutövning uppfyller kriterierna för brottet misshandel i 3:5 BrB genom att skära och skada patientens kropp. Dock betraktas läkarens ingrepp som socialt adekvata då de sker i syfte att bota eller lindra sjukdom och således har ett legitimt syfte. Dock under förutsättning att patienten har lämnat samtycke och att det sker inom ramen för den samhällsviktiga sjukhusverksamheten.⁷⁹

Även i idrottssammanhang fyller läran om social adekvans en betydelsefull funktion. Sporter som exempelvis boxning och hockey medför en acceptans av visst fysiskt våld inom ramen för idrottsutövningen. Utanför idrotten skulle dessa våldshandlingar betraktas som direkt misshandel enligt 3:5 BrB och medföra straffrättsligt ansvar. Högsta domstolen har i NJA 2018 s. 591

⁷⁵ NJA 2018 s. 591 p. 9 och p.12.

⁷⁶ NJA 2018 s. 591 p. 10; Jack Ågren, *Social adekvans*.(2021) s.12–13.

⁷⁷ Asp m.fl. (2013) s. 254–255.

⁷⁸ NJA 2018 s. 591 p. 11 och p. 14-16, P. Asp & Ulväng (2019) s. 44–45.

⁷⁹ *ibid*.

prövat gränserna för hur långt det accepterade våldet sträcker sig inom ramen för social adekvansläran. I fallet utdelar en hockeyspelare ett slag mot en annans spelares nacke, "cross-checkning" inom hockeyvärlden. HD bedömde att slaget låg utanför gränsen för accepterat våld och därför kunde inte slaget motiveras som tolerabelt med stöd av läran om social adekvans. Hockeyspelaren dömdes således för misshandel. Rättsfallet illustrerar den intresseavvägning som görs vid tillämpning av läran om social adekvans. Det vill säga avvägningen mellan handlingens allvarlighet och regelövertamp å ena sidan och dess samhällsnytta å den andra.

6 Analys

6.1 Krocken mellan black box och insiktsförmåga

En artificiell algoritm saknar det mänskliga sinnet för medvetande och kan därigenom inte bära skuld eller ansvar.⁸⁰ Detta förskjuter ansvarsfrågan till de mänskliga individerna bakom systemen: Under vilka förutsättningar kan systemutvecklare och ingenjörer hållas straffrättsligt ansvariga för skador orsakade av systemet?⁸¹

Konformitetsprincipen utmanas direkt av black box-problematiken. Principen föreskriver att endast den som hade *förmåga och tillfälle* att rätta sig efter lagen, kan hållas straffrättsligt ansvarig. De komplexa DNN-systemen och deras brist på transparens, medför att varken ingenjör eller systemutvecklare har insyn i systemets beslutsprocesser i det specifika gärningsögonblicket. De kan därför inte anses ha förmåga eller tillfälle att vidta den akt-samhet som lagen kräver. Att utkräva straffrättsligt ansvar av ingenjörerna trots denna avsaknad av insyn och kontroll, skulle medföra ett avsteg från legalitetsprincipens krav på förutsägbarhet. Detta styrks även av vad som framhålls i SOU 2018:16 där utredningen konstaterar att självkörande bilar utmanar konformitetsprincipen på grund av svårigheterna att fastställa vem som kan hållas rättsligt ansvarig.

Black box-problematiken utmanar vidare brottsbegreppets traditionella tvåstegsbedömning, vilket skapar svårigheter att uppfylla täckningsprincipen. Den objektiva sidan av brottet kan konstateras vara uppfyllt i de ögonblick som en dödlig olycka inträffar. Täckning för den subjektiva sidan däremot hindras genom black box-problematiken eftersom det inte går att fastställa den kognitiva insikten hos exempelvis ingenjören bakom systemet. I

⁸⁰ ibid; SOU 2012:17 s. 522. Där uttrycks principen: ”*inget straff utan skuld*”.

⁸¹ Harvard Business School, 'If a car can drive itself, can it make life or death decisions?', *HBS Working Knowledge*, 2024 (Jfr Badaraccos argument rörande ledarens ansvar att integrera etiska överväganden initialt i utvecklingsprocessen. I kontexten för autonoma fordon kan detta innebära att ingenjörsteamet har en utökad aktsamhetsplikt att undersöka systemets säkerhetsgränser, särskilt vid oförutsägbara AI-system.)

ljuset av NJA 2023 s.29 åskådliggör HD den höga nivå av bevis som domstolen uppställer i syfte att kunna utesluta alternativa händelseförlopp och gärningsmän. Vidare måste det kunna styrkas bortom varje rimligt tvivel vem som kan hållas ansvarig. Dessa höga krav på bevisningen blir ytterst svåra att uppnå vid AI-relaterade olyckor eftersom black box-systemens brist på transparens och insyn i händelseförloppen, snarare skapar *rimliga tvivel*. Black box-problematiken öppnar i stället upp för möjliga alternativa händelseförlopp, vilka åklagaren inte har möjlighet att bevisa på grund av bristen på transparens. Kravet på subjektiv täckning kan därigenom inte heller uppnå den rimliga standard som HD belyser genom prejudikatet.

Mot denna bakgrund landar jag i slutsatsen att varken ingenjör eller annan systemutvecklare, kan klandras på straffrättslig grund då ingen av dessa kan hållas ansvariga för en teknisk oförutsägbar händelse. Att ställa ingenjörer eller systemutvecklaren till rätta för ansvar, när dessa saknar insiktsförmåga genom systemens tekniska komplexitet, innebär ett kringgående av skuldprincipen, *nulla poena sine culpa* och skapar en godtycklig tillämpning av det straffrättsliga ansvaret. I det långa loppet riskerar detta även att omvandla skuldprincipen till ett strikt ansvar, vilket enligt mina resonemang, riskerar urholka syftet med straffrättens syn på personlig skuld.

6.2 Aktsamhetsnormen i riskfylld verksamhet

Bedömningen av huruvida ingenjörerna bakom systemen har agerat omedvetet oaktsamt blir komplex vid självkörande nivå 5-fordon. Detta kräver en tvåstegsprövning i ljuset av konformitetsprincipen. Det första steget i bedömningen av omedveten oaktsamhet, det vill säga orsaksledet, fastställs genom att bedöma vad som hade krävts av en aktsam person i likartad situation. Det vill säga; den objektiva aktsamhetsnormen. Frågan blir således vilken aktsamhetsnorm som gäller vid skapandet av autonoma fordon. NJA 2006 s.228 åskådliggör synen på aktsamhet, där HD anger att kravet på aktsamhet är högre inom en riskfylld verksamhet, med högt ställda krav på noggrannhet och särskild uppmärksamhet, exempelvis i ett sjukhus. Mot denna bakgrund anser jag att skapandet av självkörande fordon utan insyn, i

högsta grad kan betraktas som en riskfylld verksamhet och därigenom motiverar en långtgående aktsamhetsnorm som ingenjörerna måste vidta i sitt arbete. Exempelvis genom att vara särskilt uppmärksamma och noggranna vid programmeringen och utvecklingen av de självkörande bilarna. Att driftsätta självkörande fordon med en black box, utan möjlighet till noggrann insyn och uppmärksamhet över systemens beslutsprocesser, är enligt min mening att frångå denna aktsamhetsnorm och bör ses som mycket oaktsamt.

Det andra steget av bedömningen för omedveten oaktsamhet, det vill säga klandervärdhetsledet, prövar i vilken utsträckning som individen hade förmåga och tillfälle att iaktta och inse den rådande aktsamhetsnormen. Appliceringen av HD:s syn på oaktsamhet i NJA 2006 s.228 måste utifrån detta led i bedömningen, dock ske med viss försiktighet och restriktiv tolkning av HD:s synsätt. Sjuksköterskan i fallet har direkt kontroll över injektionen, medan ingenjören och systemutvecklaren skapar en algoritm som i sin tur självständigt skapar sin egen logik för inlärning oberoende av mänskliga aktörer tidigare i ledet. Ingenjörerna har således ingen direkt kontroll över algoritmens beslutsprocesser eftersom systemen agerar autonomt. Ingenjören kan följa branschstandarder korrekt och iaktta den aktuella aktsamhetsnormen, och systemet kan ändå framkalla ett oförutsägbart fel som ingenjören inte har någon kontroll över. Bedömningen i klandervärdhetsledet faller om kravet på kontroll saknas. Black box-problematiken är så pass komplex att inte ens ingenjören som skapat systemet kan få insyn eller förutse systemets nästa steg. Min slutsats blir således att det är rättsligt oförenligt med skuldprincipen att klandra en ingenjör för omedveten oaktsamhet, när ingenjören saknar de tekniska förutsättningarna som krävs för att få insikt om systemens eventuella dolda fel.

6.3 Bevisvärderingens moment 22

Inom svensk straffrätt har åklagaren det yttersta ansvaret att styrka åtalet mot den tilltalade, ställt utom rimligt tvivel, vilket är ett högt ställt beviskrav.⁸²

Högsta domstolens praxis förtydligar hur indiciebevisning ska värderas, vilket blir relevant för att belysa de höga beviskraven som HD utgår från i sin bedömning. I rättsfallet ”*mordet vid Årsta torg*” konstaterar HD att om utrymme finns för alternativa gärningsmän, eller oklarhet råder kring den faktiska inblandningen bland gärningsmännen, kan beviskravet inte anses uppfyllt. Vidare framhåller HD i rättsfallet ”*Mordet vid busshållplatsen*” att bevisningen måste inneha sådan styrka att det kan ställas bortom rimligt tvivel att den tilltalade var gärningsman och att inga alternativa gärningsmän kan styrkas.

Dessa prejudikat åskådliggör den konkreta utmaning som aktualiseras vid olyckor orsakade av självkörande bilar. Black box-problematiken osynliggör den interna logiken i systemen, vilket resulterar i bevisrättsliga hinder för åklagarsidan och begränsar domstolens underlag för en rättssäker ansvarsprövning. Det blir i efterhand tekniskt omöjligt att rekonstruera händelseförloppet kring varför algoritmerna agerade som de gjorde i gärningsögonblicket och huruvida ingenjörerna hade något ansvar för dessa fel. En åklagare kan därför få svårigheter att bevisa och utesluta alternativa tekniska fel, oberoende av ingenjörernas handlande, och saknar därigenom teknisk bevisning att styrka varför en algoritm fattade ett visst beslut. Detta utmynnar enligt mig i ett moment 22: Där systemens natur av dold logik utan insyn och transparens, gör det tekniskt omöjligt att ställa det bortom varje rimligt tvivel huruvida en ingenjör låg bakom det tekniska felet, eller om felet berodde på ett klandervärt agerande hos ingenjören. Åklagaren saknar helt möjlighet, med stöd av nuvarande lagstiftning, att kunna åtala någon mänsklig individ. Genom detta skapas oavsiktligt en form av ansvarsfrihetsgrund av bevisrättslig karaktär.

⁸² 20 kap. 2 och 6§§ samt 45 kap.1§ RB.

Åklagarsidan möter vidare svårigheter i att styrka den kognitiva insikten och därigenom omedveten oaktsamhet för de individerna bakom systemen. Konsekvensen av detta skulle kunna resultera i att ingenjörerna liksom systemutvecklarna, inte känner lika stort incitament till att iaktta den höga graden av aktsamhet, då straffrättsligt ansvar inte kan utdömas för deras fel. Ur ett större perspektiv riskerar detta skada rättssäkerheten i samhället och grundläggande principer som allas likhet inför lagen samt skuldprincipen. Dessa resonemang styrks även av Advokatsamfundets diskussion kring kravet på en klar och tydlig lagstiftning för att bortom varje rimligt tvivel och vid varje given tidpunkt, kunna fastställa vem som ska hållas straffrättsligt ansvarig.

6.4 Värdering av social adekvans som nödfallsutgång

Avslutningsvis är det av stort intresse att analysera huruvida läran om social adekvans kan användas som en rättfärdigande omständighet vid dödliga olyckor orsakade av AI-styrda bilar. Frågan är om man i analogi med resonemangen i NJA 2018 s. 591 kan undanta en ingenjör från straffrättsligt ansvar med motiveringen att handlingen skett inom spelets regler för teknisk innovation? Inom idrotten accepteras vissa handlingar som annars betraktas som brottsliga med stöd av att dessa handlingar sker inom spelets regler och således väger det samhällseliga intresset tyngre än bruket av våldshandlingar. Det är möjligt att argumentera för att samhället, genom att tillåta nivå 5-fordon, accepterar att eventuella risker för att olyckor kan inträffa, i utbyte mot en säkrare trafikmiljö på sikt och färre olyckor.⁸³

Att tillämpa läran om social adekvans i kontexten självkörande fordon kan dock medföra problematik och kritik. I kontrast till en hockeyspelare, har ett offer i trafiken inte samtyckt till den riskexponering som finns inom vissa sporter och som sportdeltagare underförstått samtycker till. Att tillämpa en

⁸³ SOU 2016:28 s.112 f.

oskriven ansvarsfrihetsgrund med det primära syftet att täppa igen det ansvarsvakuum som uppstår genom Black box-problematiken, lever inte upp till den rättssäkerhet som Advokatsamfundet eftersträvar i sitt remissvar. Ur ett rättssäkerhetsperspektiv blir det vidare problematiskt att legitimera handlingar med dödlig utgång med stöd av läran om social adekvans. En rimlig tolkning av praxis bör därför enligt mig vara att betrakta social adekvans som en otillräcklig lösning, då den annars riskerar att urholka straffrättens fundamentala krav på personlig skuld.

7 Slutsatser och rättspolitisk diskussion

7.1 Rättssäkerhet kontra det systematiska vakuumet

Mot bakgrund av det ovan anförda landar jag i slutsatsen att gällande rätt är otillräckligt utformad för att hantera de straffrättsliga utmaningar som nivå-5 fordon aktualiserar. Detta åskådliggörs särskilt genom black box-problematiken, vilken medför att det i praktiken blir omöjligt att styrka en klandervärd insikt hos ingenjörerna i enlighet med rekvisiten för omedveten oaktsamhet. Detta skapar ett tydligt ansvarsvakuum som enligt mina resonemang omöjliggör att i praktiken hålla någon mänsklig aktör straffrättsligt ansvarig vid AI-olyckor.

I likhet med vad Europarådet understryker är det av yttersta vikt med en förutsägbar och tydlig lagstiftning för att bibehålla straffrättens legitimitet samt dess upprätthållande av rättssäkerheten i samhället.⁸⁴ Lagstiftaren bör därför vidta åtgärder för att hantera konflikten mellan skuldansvar och artificiell autonom teknik. Denna intressekonflikt förutsätter en avvägning mellan individens rätt till en förutsägbar rättstillämpning å ena sidan och samhällets behov av ansvarsutkrävande strafflagstiftning å den andra.

7.2 Konflikten mellan skuldprincipen och strikt ansvar

Det rådande ansvarsvakuumet aktualiserar frågan om vilken ansvarsmodell som bör tillämpas i syfte att fylla igen vakuumet. Lundqvist argumenterar

⁸⁴ S. Gless & A. Peralta Gutiérrez, 'Discussion Paper on Criminal Liability Related to AI systems', *European Committee on Crime Problems (CDPC)*, CDPC(2024)09, 2024, s. 5 och 7; Jfr s. 7 angående konflikten mellan AI-systemens "inherent unpredictability" och kravet på legalitet samt s. 5 rörande vikten av "clear legislation" för att säkerställa förutsebarhet enligt artikel 7 EKMR.

för strikt ansvar där tillverkaren eller utvecklaren ska hållas strikt ansvarig för de AI-orsakade skadorna.

Lundqvists modell för strikt ansvar riskerar dock att urholka skuldprincipen *nulla poena sine culpa*. Att hålla en aktsam systemutvecklare ansvarig för oförutsägbara fel vore ett radikalt kringgående av den rättsfilosofiska synen på tillräknelighet och fri vilja som präglat synen på straff i århundraden. Vidare anser jag att ett införande av strikt ansvar med hög sannolikhet hade inneburit en kränkning av den enskildes rättssäkerhet. Detta då strikt ansvar riskerar att omvandla straffrätten till ett civilrättsligt sanktionsverktyg och därigenom åsidosätta konformitetsprincipens krav på individens förmåga att förstå och rätta sig efter lagen.

7.3 Transparens som lösning: återställande av skuldprincipen

En rättssäker lösning som går i linje med EU:s etiska riktlinjer på ansvarsskyldighet och transparens⁸⁵ och Advokatsamfundets syn på tydlighet, är att genom lagstiftningsvägen införa krav på Explainable AI (XAI). Denna tekniska lösning innebär att AI-systemens black box, i stället blir transparenta och därigenom omvandlas till vita lådor, vars beslutskedjor blir läsbara för människor.⁸⁶ Implementering av en sådan vit låda med krav på teknisk transparens, får en direkt positiv inverkan på det straffrättsliga området. Exempelvis skulle bevisläget förändras radikalt och ge åklagaren goda förutsättningar att rekonstruera händelseförloppet och följa den bakomliggande beslutskedjan och dess processer. Därigenom skapas även goda möjligheter att styrka subjektiv täckning och uppfylla täckningsprincipen och även utesluta alternativa tekniska fel. Vidare ges domstolen tydligare grunder att pröva varifrån bristen härstammade samt huruvida det rörde sig om en

⁸⁵ EU-kommissionen, *Ethics guidelines for trustworthy AI*, High-Level Expert Group on Artificial Intelligence, 2019, s. 16 p. 4 och 7.

⁸⁶O. Picken (red.), 'Explainable AI: Addressing Challenges for Machine Learning Implementation', *Oxford Global*, 2022, F. Hussain m.fl., 'Explainable Artificial Intelligence (XAI): An Engineering Perspective', *arXiv preprint*, 2021, s. 1–2.

oundviklig olyckshändelse eller orsakad av faktisk oaktsamhet och mänsklig brist.

Framtida implementering av XAI som krav för att tillåtas driftsätta självlärande bilar i trafiken, återställer med andra ord bevisförutsättningarna som är nödvändiga för en korrekt oaktsamhetsbedömning. Till skillnad från strikt ansvar, riskerar inte XAI att urholka straffrättens syn på skuld och bibehåller således även tilliten för rättssystemet och legitimitetsprincipens syfte. Straffrättens funktion i samhället är inte att fungera som en garanti för skadeståndsrättslig ekonomisk kompensation, utan i stället vara en garanti för att utkräva straffrättsligt ansvar mot klanderfulla och oaktsamma brottshandlingar. XAI kan visserligen få negativa konsekvenser ur en ekonomisk och konkurrenskraftig synvinkel, där endast teknikjättar har råd att efterleva och implementera XAI-kravet, vilket skulle kunna hämma den tekniska konkurrensen. Samtidigt måste rättssäkerheten enligt min bedömning väga tyngre jämfört med de marknadsmässiga nackdelar som XAI kan förelämba.

Genom ovanstående kan det avslutningsvis konstateras att det är omöjligt att straffa det oförutsägbara. Kravet på transparens blir därför avgörande; det täpper inte bara till vakuumet mellan dold teknik och mänskligt ansvar, utan säkerställer även att straffrättens syfte bibehålls i en alltmer artificiellt intelligent styrd värld.

Källförteckning

Offentligt tryck

EU-rättsliga källor

Europaparlamentets och rådets förordning (EU) 2024/1689 av den 13 juni 2024 om fastställande av harmoniserade regler för artificiell intelligens (AI-förordningen).

Svenska författningar

Brottsbalk. (1962:700).

Rättegångsbalk. (1942:740)

Propositioner

Prop. 1993/94:130 *Ändringar i brottsbalken m.m. (ansvarsfrihetsgrunder m.m.)*.

Prop. 1994/95:23 *Ett effektivare brottmålsförfarande*.

Prop. 2017/18:177 *En ny sexualbrottslagstiftning byggd på frivillighet*.

Statens offentliga utredningar

SOU 2002:3 *Psykisk störning, brott och ansvar*.

SOU 2012:17 *Psykiatri och lagen – tvångsvård, straffansvar och samhällsskydd*.

SOU 2016:28 *Vägen till självkörande fordon – försöksverksamhet*.

SOU 2018:16 *Vägen till självkörande fordon – introduktion*.

Rättspraxis

Högsta domstolen

NJA 2006 s. 228

NJA 2011 s. 563

NJA 2011 s. 725

NJA 2012 s. 564

NJA 2016 s.702

NJA 2018 s. 591

NJA 2020 s. 397

NJA 2022 s. 826

NJA 2023 s. 29

Högsta domstolens mål B 2354-22

Litteratur / Lagkommentarer

Asp, P., kommentar till 1 kap. 2 § brottsbalken (1962:700), Brottsbalken En kommentar, (JUNO, hämtad 2025-12-19).

Asp, Petter & Ulväng, Magnus, *Straffrätt: en kortfattad översikt*, 3 uppl., Iustus, Uppsala, 2019.

Asp, Petter, Ulväng, Magnus & Jareborg, Nils, *Kriminalrättens grunder*, 2 uppl., Iustus, Uppsala, 2013.

Hydén, H, Kleineman, J, m.fl., *Juridisk metodlära*, 3 uppl., Studentlitteratur, Lund, 2025.

Jareborg, Nils, *Straffrättsideologiska fragment*, Skrifter från Juridiska fakulteten i Uppsala, Iustus, 1992.

Nilsson, G., kommentar till 3 kap.7§ brottsbalken (1962:700), Lexino, (Karnov, hämtad 2026-01-01).

Tapani, Jussi & Tolvanen, Matti, *Straffrättens ansvarslära*, Norstedts Juridik, Stockholm, 2016.

Tyrell, Aine Josephine & Douglas-Jones, Ben, *AI in Criminal Law*, Bloomsbury Professional Law Insight, 2025.

Wennberg, Suzanne, *Introduktion till straffrätten*, 13 uppl., Norstedts Juridik, Stockholm, 2022.

Ågren, Jack, *Social adekvans: ansvarsfrihet utan lagstöd*, Nordstedts juridik, Stockholm, 2021.

Artiklar och rapporter

Gless, S. & Peralta Gutiérrez, A., *Discussion Paper on Criminal Liability Related to AI systems*, European Committee on Crime Problems (CDPC), CDPC(2024)09, 2024.

Hussain, Fatima m.fl., 'Explainable Artificial Intelligence (XAI): An Engineering Perspective', *arXiv preprint*, 2021.

Jareborg, N., "Rättsdogmatik som vetenskap", *SvJT*, 2004, s. 1.

Lundqvist, Ulf, 'Om betydelsen av komparativa studier för rättstillämpningen och framväxten av en global rättsordning', *SvJT*, 2019, s. 818.

Lundqvist, U., "Artificiell Intelligens — rättsordning och rättstillämpning", *SvJT*, 2020, s. 388.

Lundqvist, F. & de Richelieu, R., *How machine learning drive the development of autonomous cars – A study of emerging technologies and the evolving value chain*, Master's thesis E2019:100, Chalmers tekniska högskola, 2019.

Maastricht Institute for Criminal Sciences (MICS), *Criminal behavior and accountability of artificial intelligence systems*, Maastricht University, 2021.

Picken, Oliver (red.), 'Explainable AI: Addressing Challenges for Machine Learning Implementation', *Oxford Global*, 2022.

Utesch Fabian, m.fl., "Towards behaviour-based testing to understand the black box of autonomous cars", *European Transport Research Review*, vol. 12, nr 40, 2020.

Wakabayashi, D., 'Self-Driving Uber Car Kills Pedestrian in Arizona', *The New York Times*, 20 mars 2018.

Elektroniska källor

Agarwal, S. & De Freitas, J., "Acceptance of Automated Vehicles Is Lower for Self than Others", *Harvard Business School Working Knowledge*, 2024, <https://hbswk.hbs.edu/item/acceptance-of-automated-vehicles-is-lower-for-self-than-others> (hämtad 2026-01-01).

Brown University, "Researchers uncover similarities between human and AI learning", *Brown University News*, 4 september 2025, <https://www.brown.edu/news/2025-09-04/ai-human-learning> (hämtad 2025-12-07).

EU-kommissionen, *Etiska riktlinjer för tillförlitlig AI*, High-Level Expert Group on Artificial Intelligence, 2019, https://www.europarl.europa.eu/me-etdocs/2014_2019/plmrep/COMMITTEES/JURI/DV/2019/11-06/Ethics-guidelines-AI_SV.pdf (hämtad 2025-12-23.)

Harvard Business School, "If a car can drive itself, can it make life or death decisions?", *HBS Working Knowledge*, 2024, (hämtad 2025-12-17).

IEEE Standards Association, *The IEEE Global Initiative 2.0 on Ethics of Autonomous and Intelligent Systems*", <https://standards.ieee.org/>, (hämtad 2025-12-23).

Imagination Technologies, "Artificial Intelligence in Autonomous Vehicles", *Future of Automotive*, <https://www.imaginationtech.com/future-of-automotive/how-do-autonomous-cars-work/artificial-intelligence-in-autonomous-vehicles/> (hämtad 2026-01-01).

Johnson, R. & Cureton, A., "Kant's Moral Philosophy", *The Stanford Encyclopedia of Philosophy*, 2022, <https://plato.stanford.edu/entries/kant-moral/> (hämtad 2025-12-05).

Knight, W., "The Dark Secret at the Heart of AI", *MIT Technology Review*, 2017, <https://www.technologyreview.com/2017/04/11/106206/the-dark-secret-at-the-heart-of-ai/> (hämtad 2025-12-23)

Macleod, C., "John Stuart Mill", *The Stanford Encyclopedia of Philosophy*, 2020, <https://plato.stanford.edu/entries/mill/> (hämtad 2025-12-23).

Nyholm, S., "Unboxing the black box of AI", *Utrecht University*, 2021, <https://www.uu.nl/en/organisation/in-depth/unboxing-the-black-box-of-ai> (hämtad 2025-12-05.)

O'Connor, T. & Franklin, C., "Free Will", *The Stanford Encyclopedia of Philosophy*, 2021, <https://plato.stanford.edu/entries/freewill/> (hämtad 2025-12-23).

SAE International, "SAE Levels of Driving Automation™ Refined for Clarity and International Audience", 2021, <https://www.sae.org/news/blog/sae-levels-driving-automation-clarity-refinements>, (hämtad 2025-11-29)

SAE International, *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*, J3016_202104, 2021 (hämtad 2025-12-11).

Stanford Institute for Human-Centered AI (HAI), "Brief Definitions of Key Terms in AI", 2023, <https://hai.stanford.edu/> (hämtad 2025-12-9).

Sveriges advokatsamfund, *Remissyttrande över betänkandet Vägen till självkörande fordon – introduktion (SOU 2018:16)*, 2018/0777, 2018, (hämtad 2025-12-23).

University of Oxford, *Key points of investigation and analysis on traffic accidents involving intelligent vehicles*, 2021, (hämtad 2025-11-29)

University of Oxford, Kwiatowska Marta, *"Are we safe in self-driving cars?"*, 2017, <https://www.cs.ox.ac.uk/innovation/research-impact/case-marta-self-drive-car.html>, (hämtad 2025-12-01).