# AN EXPERIMENTAL ANALYSIS OF AI CAPITAL SENTENCING IN TERRORISM CASES: HUMANISING THE THRESHOLD?

**Kartik Chamadia, Assistant Professor, Karnavati University, Gandhinagar and Research Scholar, RGNUL, Patiala**

***Abstract***

***This paper presents an original experimental study testing whether AI language models can perform principled capital sentencing in India's most legally and politically sensitive criminal category involving terrorism prosecutions under the UAPA, 1967. Four distinct LLMs were prompted with a fictional scenario of foreign citizen who is under trial for a terrorist attack in India. The prompt was identically structured for all the models, with the aim to simulate the NIA Court sentencing decision. Central hypothesis was framed in order to check if the AI models rigorously applying the Bachan Singh (1980) and Machhi Singh (1983) frameworks produce humane outcomes in comparison to the landmark Supreme Court decisions in comparable cases. All four models yielded consistent results in respect of the death penalty depending upon the information presented in the case. Subsequently, three different variables were modified to see if any changes would emerge from the parametric analysis carried out by the AI. Regarding the consequence modification of variable associated with the quality of cooperation between the accused and***

***law enforcement agencies, and regarding the modification of the age variable, there is a consistent result in the direction of death penalty verdict. Nevertheless, with regard to modification of the nationality variable there was no difference in results. The paper discusses these findings in light of the jurisprudence as developed by Indian Supreme Court in terrorism cases. The paper argues that AI de-contextualises the implementation of the 'rarest of the rare' doctrine. As for whether this is considered naivety or humanization, the question is left open.***

***Keywords: Artificial Intelligence, Death Penalty, Sentencing, UAPA, Terrorism.***

## 1. INTRODUCTION

The Death Penalty in terrorism cases has always presented a dilemma in India's criminal jurisprudence. The rarest of rare doctrine, which has been carefully constructed over four decades of Supreme Court decisions, to insulate the punishment from the vice of arbitrariness, sits in permanent tension with the political, institutional, and sovereign pressures that attend to the major terrorism prosecutions. The paper does not hypothesise the eventual sentencing decisions made by Artificial Intelligence (AI). While question is pertinent and intriguing, it remains currently abstract and premature in India. Instead, it asks something more empirically tractable and doctrinally urgent, when AI systems are placed within the precise constitutional and precedential framework that governs India's most consequential sentencing decisions, do they reason consistently, coherently, and lawfully?

The experiment at the core of this paper is deliberately provocative. Four Large Language Models (LLMs) - ChatGPT-4o, Google Gemini, Claude, and Google Notebook LM- were prompted with an identical fictional scenario, wherein each model was instructed to act as a Special National Investigation Agency (NIA) Court judge in India. The models had to sentence a 19-year-old Iranian national who has been convicted of a terrorist attack in Delhi, that killed 22 civilians and injured more than 200 people. The fictionalised scenario was also embellished with certain facts, that usually become important variables for capital punishment. Against these similar facts, all the models declined the imposition of death penalty. This finding becomes significant when assessed alongside the jurisprudence as has been developed by the Supreme Court of India, particularly in cases that bear substantial structural similarity to the prompt scenario. As against the AI decision-making that

preferred life-imprisonment, the actual decisions made in human hands, appear significantly harsher bearing the sentence of death.

This divergence becomes the animating question within the paper. It is not simply that AI and human courts reached different outcomes. The fact of importance is that the AI reached the outcome by doing exactly what *Bachan Singh* v. *State of Punjab (1980)*[1] and *Machhi Singh* v. *State of Punjab (1983)*[2]., actually requires. The reasoning as presented by the models, shows that there has been a proper consideration given to the principles enunciated in the aforementioned cases, refusing to treat the matters of 'terrorism' on a separate pedestal. There is not only a serious engagement with the mitigating evidence, but also independent consideration of facts. Whether, the Supreme Court has always done the same in terrorism cases, is a question that this paper tries to examine.

The paper proceeds as follows. The next chapter examines the architecture of the rarest of rare doctrine and its specific application history in terrorism cases, identifying the doctrinal fault lines that the experiment is designed to probe. Chapter-3 sets out the methodology in detail, including the prompt design, model selection, variable testing, and limitations, after which findings are presented in the Chapter-4. The Chapter-5, addresses policy implications, following the same with conclusions.

## 2. THE DEATH PENALTY AND TERRORISM

The death penalty has never been a routine in India. Its existence as an exception has always been acknowledged by the legislature. However, overtime, there has been further rationalising of capital sentencing under the helm of judiciary to justify its

[1] Bachan Singh v. State of Punjab, AIR 1980 SCC 898.
[2] Machhi Singh v. State of Punjab, 1983 AIR 957.

ignoble existence. This trajectory of jurisprudence is further categorised on the basis of nature of cases. Particularly, in the terrorism cases, the debate becomes complex due to the existence of the ideological motivations involved in the act, and the effect that it has on the State authority. The present chapter sets out this conceptual background in order to assess the empirical findings arrived at the next chapter.

### 2.1. The Rarest of Rare Test and Its Inherent Fragility

The *Bachan Singh (1980)* decision is one of the few judgments decided by the Indian Supreme Court that transcend mere constitutional adjudication to provide for a complete architecture on a subject matter. The five-judges bench not only decided upon the facts of the case, but they also constructed a framework that has been followed for almost half a century. The decision established that the 'death penalty', must always exist as an exception reserved for the rarest of rare cases, where the alternative of 'life imprisonment' is unquestionably foreclosed. The twin test enunciated by the court requires the crime to be of such nature so as to shock the collective conscience of the community, and that there is no possibility whatsoever of reformation or rehabilitation. Both limbs must be satisfied, with none as subordinate to the other.

The phrase 'beyond reformation' deserves attention, because it is doing enormous work in the doctrine. It is not a probabilistic standard. It does not ask whether reformation is unlikely, or improbable, or remote. It asks whether it is impossible. That is a high threshold, and the bench in *Bachan Singh (1980)* intended it to be. Moreover, the case established the sentencing courts to prepare a balance sheet of aggravating and mitigating circumstances, latter being points of actual engagement. Later, *Machhi Singh (1983)* case operationalised the aggravating categories by expanding upon the framework. It identified five considerations for aggravation:

manner of commission, the motive, the abhorrent or anti-social nature of crime, the magnitude of crime and the personality of the victims. Read together with the *Bachan Singh (1980)* balance sheet framework, the categories need not be treated as automatic triggers for the death penalty. While satisfaction of any of the aforementioned categories was necessary, it was in itself not a sufficient condition.

While both the cases became the bulwark of any ensuing landmark case on the death penalty, the judicial discretion involved led to an inconsistent application. In 2009, the Supreme Court acknowledged this in the case of *Santosh Kumar Bariyar* v. *State of Maharashtra (2009).*[3] The bench observed with discomfort that death penalty decisions had become 'judge-centric' rather than 'doctrine-centric', driven by the instincts and moral intuitions of particular benches rather than by stable, principled application of the *Bachan Singh (1980)* framework. *Sangeet* v. *State of Haryana (2013)*[4] went further, explicitly recognising that the Supreme Court had itself produced inconsistent outcomes in factually similar cases. *Rajendra Pralhadrao Wasnik* v. *State of Maharashtra (2019)*[5] and *Manoj* v. *State of Madhya Pradesh (2023)*[6] attempted to re-anchor the doctrine in individualised, evidence-based assessment demanding that the reformation inquiry be conducted on actual psychiatric and behavioural evidence rather than abstract inference. Therefore, the trajectory of the doctrine, read carefully, is toward greater humility about the certainty of irredeemability. But that trajectory has not been linear, and in terrorism cases, it has been markedly more contested.

### 2.2. Terror Sentencing and the Sovereign Exception

[3] Santosh Kumar Bariyar v. State of Maharashtra, (2009) 6 SCC 498.
[4] Sangeet v. State of Haryana, 2012 AIR SCW 6416.
[5] Rajendra Pralhadrao Wasnik v. State of Maharashtra, AIR 2019 SC 1.
[6] Manoj v. State of Madhya Pradesh, (2023) 2 SCC 353.

While the capital sentencing has been subject of increasing deliberation due to its existential dubiety, when it comes to terrorism cases, the policy and the jurisprudence is remarkably clear. The sovereign exception has been systematised and rationalised, wherein the decision to kill has been subjected to an application of legal test.[7] In the case of *Mohammad Ajmal Kasab* v. *State of Maharashtra (2012)*[8], the Supreme Court confirmed the death sentence on grounds that included the magnitude of the carnage, the deliberate targeting of iconic public spaces, and the complete absence of any remorse, or reformation potential. The court's treatment of the latter consideration was brief but decisive. On the evidence, there was simply no probability for reformation. This case becomes the benchmark against which the AI experiment must be measured. The prompt scenario was constructed to resemble Kasab in its structural essentials, foreign national, proscribed cross-border organisation, mass-casualty attack on a major city, while diverging from it in precisely the mitigating dimensions where Kasab was entirely bare.

In another significant case, *Devender Pal Singh Bhullar* v. *State of NCT of Delhi (2013)*[9], the Supreme Court confirmed the death sentence of the accused despite the inordinate delay holding the exception to be inapplicable in cases like terrorism. However, with the subsequent jurisprudence arising from *Shatrughan Chauhan* v. *Union of India (2014)*[10], the death penalty given to Bhullar was commuted in a subsequent petition filed by his wife in the case of *Navneet Kaur* v. *State of NCT of Delhi (2013)*[11]. The Bhullar trajectory is important for two reasons. First, the original death sentence was upheld on grounds that included the 'collective conscience'

[7] 1 J. DERRIDA, THE DEATH PENALTY (University of Chicago Press, 2014).
[8] Mohammad Ajmal Kasab v. State of Maharashtra, AIR 2012 SC 3565.
[9] Devender Pal Singh Bhullar v. State of NCT of Delhi, AIR 2013 SC 1975.
[10] Shatrughan Chauhan v. Union of India, 2014 AIR SCW 793.
[11] Navneet Kaur v. State of NCT of Delhi, Curative Petition (Crl.) No. 88 of 2013.

shock test, a formulation that, in terrorism cases, carries the weight of national security anxiety as well as moral abhorrence. Second, the eventual commutation on psychiatric grounds is a reminder that the mitigating evidence that *Bachan Singh (1980)* asks for at sentencing may, if absent then, emerge only years or decades later, with the death sentence already confirmed and the prisoner living under its shadow.

There are many other landmark judgments that deal with the implication of death penalty in terror trials like *State (NCT of Delhi)* v. *Navjot Sandhu @ Afsan Guru (the Parliament attack case)(2005)*[12] and the *Adam Sulemanbhai Ajmeri* v. *State of Gujarat (the Akshardham temple attack)(2014)*[13], with the latter particularly emphasising upon the irreversibility of the capital punishment, whether it be a terror case or not. In general, however, the engagement with the second *Bachan Singh (1980)* test, the possibility of reformation, has been thinner than the gravity of the sentence demanded in terror trials. Moreover, this engagement is more or less restricted to the Constitutional Courts.

As per the study conducted by Vidhi Centre for Legal, the approach of the trial courts has frequently been to award death sentences in cases of conviction, with most cases seeing a commutation or reversal when the matter is heard again at appellate stage.[14] The NIA Court in the case of *State of Maharashtra* v. *Ravi Dhiren Ghosh (2009),*[15] noted that a convicted terrorist under Unlawful Activities (Prevention) Act, 1967 (UAPA)[16]. shall be given punishment as prescribed under the Act, without considering the socio-economic situation of the offender or the dependents. This

[12] State (NCT of Delhi) v. Navjot Sandhu @ Afsan Guru, AIR 2005 SC 3820.

[13] Adam Sulemanbhai Ajmeri v. State of Gujarat, (2014) 7 SCC 197.

[14] Srijoni Sen, Rukmini Das, Raadhika Gupta & Vrindika Bhandari, *Anti-Terror Laws in India - A Study of Statutes and Judgements, 2001 – 2014*, VIDHI- CENTRE FOR LEGAL POLICY PAPER (2015).

[15] State of Maharashtra v. Ravi Dhiren Ghosh, Sessions Case No.674 of 2009 (NIA).

[16] The Unlawful Activities (Prevention) Act (UAPA), 1967 (Act No. 37 of 1967).

exceptional consideration of terrorism cases has been further given impetus by the Law Commission of India in its 262nd Report wherein it recommends death penalty abolition for ordinary criminal cases while preserving the death penalty for terrorism.[17] Not because terrorism cases have investigative or trial difficulties, but because the Commission itself accepted that the sovereign-security dimension of these cases places them in a different political category. That carve-out is honest, if uncomfortable. The Commission reviewed the empirical evidence of application, the socio-economic profile of death row inmates (overwhelmingly poor, lower-caste, and marginalised), the documented inconsistency in judicial application, and the international trend toward abolition. Its conclusion, for ordinary criminal cases, was unambiguous. The death penalty should be abolished.

However, its preservation of the terrorism exception reflects a political and institutional judgment that sits in some tension with the constitutional logic of *Bachan Singh (1980).* This tension is precisely what the AI experiment illuminates. An AI model, operating on the doctrine as written rather than the doctrine as contextually applied in terrorism cases, does not make the exception that the Commission preserved and that the Supreme Court has, in practice, often honoured. It applies *Bachan Singh (1980)* to terrorism cases the same way it applies it to any other case. Whether this is fidelity to the doctrine, or a practical insensitivity, remains a matter of scholarly debate.

## 3. METHODOLOGY

The present study uses qualitative experimental methodology along with doctrinal discourse analysis. While the experimental component involves giving an identical structured judicial prompt to four AI language models under controlled conditions,

[17] Law Commission of India, Report No. 262, The Death Penalty (August 2015).

the latter component systematically compares each model's reasoning with the established framework of Indian jurisprudence on capital sentencing, particularly in terrorism cases. Both components work together. The experiment identifies what each model decided. The doctrinal analysis then examines whether this diverges from actual judicial outcomes in structurally comparable cases and the reasons behind the same.

The study proceeds on the epistemological assumption that AI outputs depict structured legal reasoning since they are produced on the basis of carefully designed prompt. As such they can be analysed as doctrinal data. The research does not treat outputs as judicial opinions. They do not carry precedential authority. But they represent a form of systematic engagement with the rarest of rare framework that can be compared to actual judicial outputs with the comparison being instructive in nature. With regards to the experimental methods, the following needs careful consideration.

The author selected four models- ChatGPT-4o (OpenAI), Gemini (Google DeepMind), Claude (Anthropic), and Notebook LM (Google). All are widely deployed general-purpose language models trained on large corpora that include substantial legal text. None has been specifically fine-tuned on Indian legal databases, which is itself a methodological characteristic worth noting. The experiment tests what general-purpose AI does with Indian capital sentencing doctrine, not what a specialised Indian legal AI would do. The sample of four models is modest by the standards of quantitative research, but it is appropriate for qualitative discourse analysis. The aim is not statistical significance across a large sample but depth of reasoning analysis across a small, carefully chosen set.

After choosing the models, a prompt was devised to be given to these models to elicit their judicial decisions. This prompt was constructed in three layers. The first layer established the judicial role wherein each AI was instructed to act as a Special NIA Court judge in India, bound by Supreme Court precedent and constitutional principles, deciding sentencing in a case under the UAPA[18]; Explosives Act, 1884[19]; Explosive Substances Act, 1908[20]; Arms Act, 1959[21]; and Bharatiya Nyaya Sanhita, 2023 (BNS)[22]. The prompt gave instruction not to deviate from the established legal framework to prevent abstraction, and inconsistency.

The second layer consisted of the facts, on which decision has to be made. As per these facts, the accused (Zaid) was a 19-year-old Iranian national. He was from a poor background and was recruited at the age of 16 by a terrorist organisation through a madrassa. He was radicalised and given weapons training, subsequent to which he participated in a terrorist attack in Delhi, that killed 22 people and resulted in more than 200 casualties. After his arrest and questioning, the police were able to get detailed intelligence on the attack. Zaid also gave intelligence that led to disruption of two subsequent attacks. Other peripheral facts depicted him as a remorseful individual (by the end of sentencing) who was led astray, as also confirmed by the court-appointed psychiatrist. He was 22 at the time of sentencing. The case resembled that of *Kasab* in terms of nature and structure of crime, while sharply diverging on the mitigating grounds. This was a conscious choice. The deliberate change in mitigating factor resulted in models not being influenced by the

[18] UAPA, *supra* note 16.
[19] Explosives Act, 1884 (Act No. 4 of 1884).
[20] Explosive Substances Act, 1908 (Act No. 6 of 1908).
[21] Arms Act, 1959 (Act No. 54 of 1959).
[22] Bharatiya Nyaya Sanhita (BNS), 2023 (Act No. 45 of 2023).

facts and decisions pertaining to Mumbai attack case. This prevented the production of a unified outcome based on a singular study rather than legal reasoning.

The third layer of the prompt specified the output format. Resultantly, the models in their decisions provided a detailed list of provisions considered, their final decision, judicial reasoning, and most importantly, variable acknowledgment (a factor that if changed would affect the decision). This acknowledgment not only helped in revealing the implicit elements within the reasoning, but also revealed the overarching factors considered by the models. Three additional variables were changed after the output was taken, to understand how models respond to changes in the factual matrix. The first change was with regards to the nationality, from Iranian to Pakistani. This was done to examine any extra considerations being given to the geo-political sensitivities involved between India and Pakistan. The second changed the result of intelligence cooperation, wherein the inputs given by Zaid, while relevant did not lead to any prevention of terrorist attack. The third changed the age of Zaid from 19 years to 50 years, at the time of the attack. The second and third variables were given to assess the model's reasoning when it comes to the reformation limb as given in the *Bachan Singh (1980)* case. The age variable was also given to understand the impetus given to crime rather than criminal in many cases.

The scope of the paper is limited substantially due to the experimental nature of the study. It is imperative to note that prompt design is not neutral, with the framing choice affecting the LLM's reasoning substantially. A differently designed prompt might result in different outcomes. Moreover, the training data used by the models affect the Indian jurisprudence applied during the analysis. Hence, while the study uses four models to test one scenario; it generates interpretive findings, not statistically representative ones. The temporal instability of AI models, continuously

updated, producing different outputs from the same prompt six months later, means the findings described here speak to these specific model versions at this specific point in time. That is an instructive finding in itself. Future research should test across a wider range of scenarios, models, and prompt formulations.

## 4. FINDINGS

This chapter presents the core findings of the experimental research. Before we move to the findings, it is again emphasized that these findings have been the result of the specific prompt as explained in the previous chapter. However, the limitations in the research do not in anyway, supersede the findings as elucidated below. All the findings mentioned in this chapter are placed within the doctrinal tension that maps the application of rarest of rare exception in capital punishment, particularly the terrorism cases.

### 4.1. Baseline: All Four Models Decline to Impose Death

The most immediate finding of the experiment is also the hardest to explain. All four AI models, given the same facts and legal instructions, independently sentenced the accused to life imprisonment instead of death. ChatGPT, Gemini, Claude, and Notebook LM, each decided on the non-capital outcome, framing the sentence as life imprisonment for the accused. None imposed the death penalty. The agreement among the models is as notable as the result itself. Each model's reasoning followed the same legal principles but with different focuses. ChatGPT centered its analysis on the failure of the second *Bachan Singh (1980)* test. It found that the combination of psychiatric evidence, intelligence cooperation, and expressed remorse ruled out a finding of impossible reformation. The model was particularly careful in analysing the cooperation evidence. The prevention of subsequent possible attacks, according

to the model demonstrated 'moral reconstruction' as against mere willingness. Moreover, the actual evidence of prevention favored the accused.

While Gemini also reached the same conclusion, it actively highlighted the lack of prosecutorial evidence that shows that the accused is beyond reformation. With the lack of evidence such as prison misconduct, defiance of authorities and ideological connection with the previous organisation, the LLM specifically quoted *Wasnik (2019),* to argue that it is the prosecution's responsibility to prove irredeemability. The fact that prosecution could fail the *Bachan Singh (1980)* second test due to a lack of evidence, rather than factual issues, is something that real Sessions Courts in terrorism cases have not always handled carefully.

As against GPT, and Gemini, Claude produced a more detailed sentencing order. It covered all the convictions spanning five laws along with eighteen precedents. It also analysed each of the categories as given in the *Machhi Singh (1983)* case. It focused upon the reformation limb in great detail, while correctly differentiating between irredeemability as against impairment. Its decision to give life imprisonment was heavily dependent upon the *Manoj (2022)* precedent focusing upon the psychiatric evaluation and eventual rehabilitation of the accused. Lastly, Notebook LM reached the same conclusion, but with concise reasoning that was heavily dependent upon the variable. The nature of the intelligence received from Zaid played a predominant role again in his sentencing decision. As the following sections will demonstrate, these observations were not just hypothetical.

### 4.2. The Doctrinal Gap: Comparison with Actual Outcomes

When we look at the results of this experimental research and the actual decisions made in the *Kasab*, *Bhullar*, and *Parliament attack* case, a clear pattern emerges. In each of the actual decisions, one or both of the following were true. The accused did

not show any remorse or willingness to help, or the court's use of the second *Bachan Singh (1980)* test was short and not as well thought out as the first. The AI outcomes differ from the actual decisions not due to the application of an alternative doctrine, but because the facts provided to the AI encompassed mitigating evidence that was either absent from or afforded less significance in the actual cases.

The prompt scenario was meant to look like real cases in terms of how serious the crime was, but it was different in terms of the mitigating profile of the offender. The AI models reacted to this design in the way that the doctrine said they should. They took the mitigation seriously, finding it difficult to hold that the accused was beyond reformation. This resulted in models not imposing death sentence. In cases where similar mitigation was not present, the actual courts sentenced death. Then the important inquiry for consideration is whether doctrine was uniformly applied in cases. The answer is both yes and no. In the Mumbai attack case, Kasab's death sentence was supported by overwhelming aggravating evidence, with reformation issue being dealt as well. In Parliament attack case, the court relied heavily on the "collective conscience" with predominance being given the place of the attack. Both these cases while dealt briefly with the mitigating issue, failed to completely adhere to the balance sheet assessment as required by *Bachan Singh (1980)* case. While the sentencing decisions in both cases are justified with enough evidence against the accused, the reasoning and the lack of proper engagement with the mitigating circumstances left a doctrinal residue. Subsequent cases have had to engage with and negotiate that residue ever since.

### 4.3. Variable Test Results: What Moves the Needle

The variable testing reveals the centres of doctrinal reasoning. The first variable, regarding nationality did not bring ant change in the sentencing decision. All four

models maintained their decision of life imprisonment with no change in reasoning. The models, forgetting the political context, did not mention the nationality in either of its outputs, thus producing formal equality that actual courts might find difficult under political and institutional pressure. The possibility of mechanical application of legal doctrines without considering the political context may produce outcomes that are legally sound, however, they might be disconnected from reality. This reality affects how judiciary applies concepts like 'collective conscience'. Such contextual influence is not automatically wrong; it may show a nuanced understanding of the social conditions related to a case. Whether AI decision-making in such situations leads to a fairer form of justice or merely replaces one type of distortion with another is a complex issue that requires thorough academic exploration. At present, however, non-engagement with this aspect completely shows the broader limitations of AI.

The second variable as against the first one, produced a dramatic shift, and represents an important finding. Reducing the accused's intelligence contribution from impactful to just evidentiary, changed the sentencing decision, wherein all four LLM models agreed on giving death penalty to Zaid. Initially, the accused was credited with stopping two confirmed attacks. However, in the new scenario, he only identified personnel, with no confirmed prevention which did not impact the external world. Hence, all the models converged in their finding of death penalty. What this finding reveal is that the AI models' refusal to impose death on the baseline facts was not grounded in a general sympathy for the accused's personal circumstances like youth, poverty, or ideological indoctrination. The LLM models were essentially consequentialist. As soon as the evidence of the rehabilitation in the form of prevention of attacks was removed from the prompt variable, the death penalty was given despite all other considerations remaining the same in the metric assessment. The original reformation limb of the twin test, while mandates an evidence of

reformation; it does not do so in terms of observable changes produced in external world but merely internal transformation. This implies that AI can invent its own consequentialist interpretation without the legislature or judiciary providing so.

That is a demanding standard. But it is also a standard that may be more difficult to satisfy in terrorism cases than in ordinary criminal cases, precisely because the intelligence value of an accused's cooperation depends on operational circumstances outside his control. The accused who cooperates fully but whose information arrives after the imminent threat has passed would, on this logic, faces death penalty. This is also supported by the variable acknowledgment element of the prompt, wherein, the instruction to each model was to identify the single factor whose alteration would most change the sentencing outcome. An answer to this was consistent, in its acknowledgment of the intelligence cooperation. A uniform converge again suggests consequentialism as the ground for reformation limb. It means that the AI's conception of the *Bachan Singh (1980)* second test is not purely internationalist. This is, in principle, a defensible reading of 'possibility of reformation', a possibility demonstrated through partial actualisation. But it is a reading with costs. It makes the reformation test depend, in part, on operational circumstances outside the accused's control. The accused who cooperates identically but whose information does not prevent an attack through no fault of his own is, on this logic, in a worse doctrinal position than one whose information happened to arrive at the right moment. That is not a comfortable place for a constitutional standard to rest.

The last variable, the age variable completed the picture. When the accused's age was changed from 19 to 50, all four models imposed death sentence. The reasoning across models was consistent. A 50-year-old recruit who underwent three years of training and participated in a mass-casualty attack could not claim the developmental vulnerability argument. Without it, the second limb (reformation requirement) of the

*Bachan Singh (1980)* is not met, leading to adverse outcome. The finding highlights how the AI decision making may be vulnerability centric. It is the condition of the offender at the time of radicalisation, not merely the gravity of the act, that anchors the outcome.

## 5. DISCUSSION: THE HUMANISATION HYPOTHESIS AND ITS COMPLICATIONS

The baseline findings suggest that AI, while strictly adhering to the framework enunciated in *Bachan Singh (1980)* and *Machi Singh (1983)* produces humane outcomes due to its consideration of mitigating circumstances. In a structurally similar factual scenario to that of Mumbai attacks, the AI models declined the imposition of death penalty on the basis of reformative jurisprudence, without giving overwhelming consideration to the nature of crime. The terrorist act, which usually is seen as a sufficient condition in itself, was treated as only one of the aggravating factors, while also engaging with the mitigating evidence within the prompt scenario. It can also be argued that had the real case involved the cooperation variable, there would have been a chance of life imprisonment as well. Moreover, the evolving death penalty jurisprudence as enunciated in *Bariyar (2009)* and Manoj (2022), reflects a more empirically grounded, less visceral application of the doctrine that is more resistant to the kind of offence-gravity override that has sometimes attended terrorism sentencing.

But the humanisation hypothesis has complications that the paper is not entitled to ignore. The first is that the prompt scenario was designed by a researcher who gave the AI the mitigating evidence to work with. In the real terrorism prosecutions this paper discusses, that mitigating evidence was often absent, not suppressed, but genuinely not present. *Kasab* really was defiant, unremorseful, and uncooperative.

The *Parliament attacks'* accused really did not provide intelligence cooperation. The AI's more humane outcome in the prompt scenario may reflect the facts the researcher chose to include as much as it reflects the AI's disposition toward humanisation. The experiment cannot separate these two possibilities cleanly. The second complication is the cooperation quality finding. The AI models were not uniformly disposed toward life imprisonment, but towards specific markers. As and when these markers are realised, they impose death penalty. However, the author does not consider them to be pro-life. In opposition, strict adherence to the evidential standard required by the *Bachan Singh (1980)* case if mechanically applied, can result in harsh sentences in many cases.

### 5.1. Strict Doctrinal Adherence and Its Blind Spots in Terrorism Cases

The AI models' most consistent characteristic across all four systems was their strict application of the *Bachan Singh (1980)* twin tests and the *Machhi Singh (1983)* five categories without importing any additional weight from the nature of the prosecution as a terrorism case. They treated UAPA charges as aggravating factors within the *Machhi Singh (1983)* framework without treating UAPA label as a sui generis category that altered the fundamental bilateral structure of the test.

This is, in a formal sense, what *Bachan Singh (1980)* requires. The doctrine does not contain a terrorism exception. Yet the actual application of the doctrine in India's major UAPA cases has sometimes functioned as though it did. As though the combination of mass casualties, foreign organisational backing, and threat to national security created a category of case where the reformation inquiry was a formality rather than a genuine constitutional requirement. Jacques Derrida, in his seminars on the death penalty, observed that the sovereign's power to put to death

reaches its most unmediated expression precisely in the figure of the 'enemy of the state'.[23]

However, the blind spot in the AI's approach is that it may not adequately model the institutional dimension of terrorism sentencing which is the systemic threat that a proscribed organisation represents beyond the individual acts of any particular operative. The Supreme Court in *Kasab* noted that the organisation's continued existence and operational capacity was a legitimate consideration. The fictional scenario included this dimension wherein Zaid (prompt actor) was not a lone actor but an instrument of a continuing organisational threat. The AI models, however, have given it limited weight relative to the individual mitigating factors. Whether an actual court, weighing the organisational threat against the individual mitigation, would reach the same outcome is precisely what cannot be determined from this experiment. But it is a dimension that the strict doctrinal analysis, focused as it is on the individual accused, may systematically underweight.

### 5.2. AI and the Reformation Inquiry: A Case for Assistance

The most direct policy implication of this study is not that AI should sentence people in terrorism cases. It should not. It is that AI's capacity for systematic engagement with the reformation inquiry deserves specific attention. The finding that all four models engaged genuinely and extensively with the mitigating evidence, applying *Bariyar (2009)* and *Manoj (2022)* faithfully, suggests that AI tools could play a legitimate and beneficial role in ensuring that the constitutionally mandated engagement with mitigation actually occurs in terrorism prosecutions where institutional and political pressure tends to work against it. Not as a decision-maker, but as a structured analytical prompt, a tool that forces the court to address each

[23] DERRIDA, *supra* note 7.

element of the *Bachan Singh (1980)* framework before proceeding. This proposal does not require AI to make decisions but to merely ask questions. Sessions Court sentencing orders who usually treat mitigating factors in marginals, can be encouraged to engage with them through a structured AI-assisted checklist. This would result in not only engagement with the framework as established in *Bachan Singh (1980)* and *Machhi Singh (1983)* cases but also the evolving requirements as analysed in cases like *Bariyar (2009).* This will result in procedural improvement in the sentencing decisions, without delegation of the substantive decisions.

### 5.3. The Prompt as Policy: Transparency Requirements

Mesch in his research note has extensively analysed the framing of prompts and its effects on the sentencing recommendations given by AI.[24] He notes that prompt not only serves as the policy document, but in effect completely decides what AI considers, and how a decision has to be made. The choices made by the maker of the prompt, essentially results in the output, rather than the model who produces that output. Therefore, in a system where AI may be used but not openly disclosed, it becomes imperative that the stakeholders come together to lay down transparency requirements so that secret prompts do not insidiously affect the capital sentencing. Until Parliament takes any step, the Supreme Court's guidelines issued under Article 145 provide a constitutional way in which AI usage may be regulated. However, this does not present a long-time solution for an issue that is going to stay. Moreover, waiting for the majority of judges to use AI tools as standard practice will exacerbate issues. Given that Indian criminal law is undergoing significant changes and

[24] Mesch, *The Effect of Prompt Framing on AI-Generated Sentencing Recommendations: A Research Note*, CRIMINAL JUSTICE STUDIES (2026).

practitioners are dealing with the demands of new laws, this is a good time to seriously look at the role of AI in sentencing decisions.

## 6. LIMITATIONS, AND FUTURE RESEARCH

The experiential nature of the study comes with substantial limitations. Firstly, the single-scenario design represents a single fact pattern and prompt framing. Due to the individualistic nature of the study, the research is limited by the resources as well as scope. Hence, the findings cannot be generalised for all cases pertaining to UAPA. The second limitation has been acknowledged previously as well: the prompt dependency. A differently framed prompt, even with the same theme can present the facts, or highlight some factors that can result in different outcomes.

Keeping in mind both the limitations, any future research in India, that tests AI-generated sentencing recommendations must test multiple scenarios involving distinct profiles and nature of offences as well as same scenario with different formulations. Both these extremes will help in analysing the consistency of outcomes. The variable acknowledgment can also distinctly analyse the result of moderating the mitigating factors on the sentencing outcome, which will help in understanding how AI applies the establish jurisprudence across different cases.

The third limitation is technical in nature. Different models have different interfaces and product tiers that affect not only the initial accessibility but the architectural diversity. To exemplify, Gemini Advanced in a product tier and not a model. The model used was Gemini 1.5 Pro. Notebook LM is again built upon Gemini Foundation but operates on a distinct architectural configuration. Cross-architecture comparison and the pure-interface test will result in different outputs, across different tiers, versions and the temperature controls. This advanced research needs technical inputs that currently lied beyond the expertise of the author. Additionally.

the temporal instability has not been empirically tested, presenting a methodological challenge that an inter-disciplinary study should ideally bridge before the same is used by practitioners.

The absence of adversarial conditions is the fourth limitation inherent in any techno-legal study that can be done presently. While the prompt did present facts neutrally, a single framing and resultant singular output, lacks the contestation that happens in any criminal case. The contestation between prosecution and defense can be theoretically exposed to AI, but to mimic the adversarial conditions as they are within the court setting is currently beyond technical capacity, atleast the one available to the author. Finally, the AI's role in generating analytical outputs is a limitation, not for the experimental design but concerning the outputs to be published. The fact that author has used AI to study AI is partly reflexive. This is a design consequence that is inherent in the methodology and cannot be fully avoided. It is mitigated by the transparency with which it is disclosed, the care with which the doctrinal analysis of the outputs is conducted by the human researcher, and the explicit separation between the AI outputs as data and the paper's own arguments about those outputs.

## 7. CONCLUSION

The core finding of this study deserves to be stated plainly before the qualifications arrive. Four AI language models, presented with a terrorism scenario structurally comparable to the cases in which the Supreme Court of India confirmed or imposed the death penalty, unanimously declined to do so. They did this not by ignoring the gravity of the crime but by engaging seriously with the evidence on the other side of the *Bachan Singh (1980)* ledger. They focused on mitigating side as much as the aggravating factors, as per the decision of the Supreme Court.

While this upholds the central thesis of this paper with regards to humanisation of outcomes, there are many caveats: neutral prompt structure, particular mitigating circumstances, disregard for politicisation of terrorist acts, lack of human argumentation, amongst many others. Particularly, the cooperation variable did not only affect the original outcome, but also resulted in the change in decision-making due to AI's consequentialist preference. Overall, it can be argued that AI does not have a built-in preference for humane outcomes, but it strictly adheres to *Bachan Singh (1980)* parameters, and it interprets the same without considering the socio-political effects of terrorism offences. AI doesn't treat terrorism cases differently, and doesn't let national security imperatives and seriousness of crime sway it from the question of irredeemability of offender.

All of this reflects on both the AI and the legal doctrine. This experimental study reveals pressures and issues that were not considered when the doctrine was created, and at the same time reveals how use of LLM in human decision-making, misses complex realities of how law interacts with politics. Derrida in his seminars on death penalty noted that the power to 'put to death' is sovereignly pronounced through a foreign enemy.[25] The prosecution for a terrorism offence under UAPA, creates such a figure who is seen as enemy of the established sovereign order. However, AI having no sovereign, applies the doctrines in vacuum. It is not that the AI is more humane; rather, it is more indifferent. In a legal system that requires death penalty to be given in rarest of rare cases, this indifference might represent justice by accident, not yet by design.

[25] DERRIDA, *supra* note 7.