IS JUSTICE JUST DATA? THE LEGAL DYSTOPIA OF PSYCHO PASS AND OUR OWN FUTURE OF LAW.

Jack Lee Hong Bo*

Abstract

This paper explores the dystopian anime *Psycho-Pass*[1] as a prescient critique of modern legal and technological developments, particularly the rise of artificial intelligence in criminal justice. At the heart of *Psycho-Pass* is the Sibyl System, an all-seeing AI that assigns citizens a "Crime Coefficient" to pre-emptively judge their likelihood of committing a crime.[2] Those who surpass the threshold are neutralized, regardless of whether they have broken the law.[3] While fictional, this predictive model has alarming parallels with real-world technologies such as COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)[4], PredPol (Predictive Policing)[5], and mass surveillance systems[6].

The paper analyses how predictive policing, algorithmic sentencing, and automated surveillance increasingly shape legal outcomes, often at the cost of due process, transparency, and human judgment. It questions whether legality grounded in statistical risk can ever align with moral justice, and whether society is trading autonomy and fairness for utilitarian efficiency. Drawing from legal theory, ethics, and real-world examples, this paper questions the legitimacy of systems that criminalize thought, suppress dissent, and automate punishment under the guise of order.

By comparing *Psycho-Pass*'s chilling depiction of AI governance with current trends in law enforcement, surveillance, and legal AI, this paper warns of a future

* LLB (Honours) candidate under the UKDTP programme, HELP University.

[1] *Psycho-Pass* (Production I.G, Fuji TV 2012–2013).

[2] *Psycho-Pass*, Season 1 Episode 1 'Crime Coefficient' (Production I.G, Fuji TV 2012).

[3] *Ibid.*

[4] Julia Angwin and others, 'Machine Bias' (*ProPublica*, 23 May 2016) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> accessed 4 August 2025.

[5] Kristian Lum and William Isaac, 'To Predict and Serve?' (2016) 13(5) Significance 14.

[6] In this article, mass surveillance systems refer to systems using real-time data to survey an area and to conduct policing. Andrew Guthrie Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement* (NYU Press 2017).

where the law becomes code and justice becomes data. It urges students, scholars, and citizens alike to critically examine how technology is reshaping legal norms, and whether we're heading toward a world where our thoughts, not just our actions, put us on trial.

From Fiction to Reality: Law in the Age of AI

The convergence of artificial intelligence and criminal justice has sparked debates across law, ethics and technology. Fictional works are not mere escapism, but speculative sites where such debates can be staged and examined.[7]

In the anime *Psycho-Pass*, a legal dystopia unfolds under the authority of the Sibyl System, an AI that continuously monitors an individual's psychological data to determine their likelihood of committing crimes. Citizens whose 'Crime Coefficients' exceed an invisible threshold are deemed threats to society and neutralized pre-emptively, despite whether or not they have actually broken the law.[8]

Though fictional, this scenario captures pressing anxieties about the rise of artificial intelligence in law enforcement, mirroring real-world developments such as COMPAS[9], PredPol[10], and biometric surveillance[11]. These technologies similarly reduce individuals to data points, generating risk scores that often evade contestation or transparency.[12]

As Thomas Giddens argued, legal fiction is not merely escapist narrative but a potent source of critique[13], stating that comics[14] are potentially significant for legal study as their fusion of words and images[15] allows for emotional engagement and further analysis with different ways of understanding knowledge.[16] Fictional works like *Psycho-Pass* thus become a vital lens to observe the trajectory of justice systems being increasingly governed by data,

[7] Thomas Giddens, 'Comics, Law, and Aesthetics: Towards the Use of Graphic Fiction in Legal Studies' (2012) 6(1) Law and Humanities 85.
[8] Psycho-Pass (n 3).
[9] Angwin and others (n 5).
[10] Lum and Isaac (n 6).
[11] Ferguson (n 7).
[12] Ferguson (n 7).
[13] Giddens (n 8) 85.
[14] *Ibid.* Comics, in this context, also includes Japanese manga, which is often associated with anime.
[15] Giddens (n 8) 88.
[16] Giddens (n 8) 100.

providing "essential human insights lost to formal thought"[17], interrogating moral inadequacies that doctrine conceals[18] through the interpretive openness and irony of fiction.[19]

Giddens' points are compelling, because it illustrates precisely why we should treat fictional works such as *Psycho-Pass* as not mere entertainment, but as a jurisprudential warning. As Oscar Wilde famously remarked, "life imitates art far more than art imitates life."[20] Fictional narratives, through compelling stories and characters, help us grasp the central messages and moral dilemmas the author presents, and encourages us to apply those lessons to our own world in ways that doctrine alone often fails to achieve.[21] Learning from examples in fiction spares us the consequences of learning only from history, and in this sense, fiction is preventive, enabling us to imagine failures of law before they manifest in practice. History shows that societies often ignore warnings until crises strike. The Titanic disaster of 1912 exemplifies this tendency, only after the tragedy did stricter maritime safety rules, such as mandatory lifeboat requirements, come into force.[22] Similarly, if policymakers and scholars treat speculative fiction seriously, works like *Psycho-Pass* may help us anticipate the dangers of algorithmic governance before they become entrenched in our courts and prisons.

Psycho-Pass warns us that as law turns into data and code, it risks legality being severed from legitimacy. When opaque algorithms replace transparent processes, authority is obeyed out of fear rather than respect, undermining the rule of law.[23] In reducing justice to numerical outputs, predictive systems strip individuals of dignity and agency, echoing critiques of bias in tools like COMPAS[24]. In Fuller's terms, predictive legality erodes the "inner morality of law," which depends upon principles such as clarity, consistency, and the ability of citizens to orient their behaviour according to known legal standards.[25] An algorithm that pre-judges individuals based on opaque statistical inferences cannot possibly meet these requirements. Fiction thus reminds us, in line with Fuller's notion of law's inner morality, that legality without legitimacy is mere control, not justice.[26]

[17] Melanie Williams, *Secrets and Laws Collected Essays in Law, Lives, and Literature* (UCL Press 2005) 38.
[18] Desmond Manderson, *Kangaroo Courts and the Rule of Law: The Legacy of Modernism* (Routledge 2012) 2, 17.
[19] *Ibid.*
[20] Oscar Wilde, *The Decay of Lying – An Observation* in *Intentions* (1st edn, James R Osgood McIlvaine and Co 1891) 1.
[21] Manderson (n 19).
[22] John P. Eaton and Charles A. Haas, *Titanic: Triumph and Tragedy, 1st edn* (W W Norton & Company 1986).
[23] Ferguson (n 7) 10. Ferguson introduces the concept of "black data" as opaque, racially coded, and distorting, which "creates legal shadows and constitutional gaps where law used to see clearly".
[24] Angwin and others (n 5).
[25] Lon L Fuller, *The Morality of Law* (rev edn, Yale University Press 1969).
[26] *Ibid.*

The Sibyl System: A Fictional Framework of Predictive Justice

In *Psycho-Pass*, the Sibyl System is an all-seeing, all-powerful AI that governs future Japan. Each member of society is assigned a Crime Coefficient score, generated by Sibyl through the evaluation of the individuals' mental states using real-time surveillance and biometric analysis. Once an individual exceeds a fixed threshold, they are labelled "latent criminals" and are either detained or executed, even without having committed any criminal act.[27]

Sibyl embodies a complete inversion of the core legal principles and safeguards of criminal law. There is no due process, no requirement of actus reus, and no presumption of innocence. Such predictive governance shifts justice from moral uncertainty to statistical logic in decision-making.[28] When an algorithmically generated risk score is available to use in law[29], it introduces "a notion of statistical prediction to a situation which was dominated by fundamental uncertainty about the outcome before"[30], inducing "a process of rewriting the temporality of the judged individual, flattening their life into a present inescapably doomed by its past"[31]. Framing justice itself as a risk management process, rather than a moral adjudication, can turn 'justice', into discrimination among different populations[32]. By collapsing law into predictive governance, Sibyl shifts justice from moral uncertainty into statistical logic, where guilt is a matter of probability rather than proof. In short, individuals under predictive governance are judged not for acts they have committed, but for the potential that an algorithm calculates they might commit crime.

The Sibyl System of our world has not yet been created, and even so, many debates would need to happen to justify its use, even in the smallest aspects of social life. However, it may be argued that at the rate that AI technology is evolving, and the normalisation of AI in our everyday life, we are moving towards a future in which AI has a say in what we would eventually do, say or even think.

This is deeply disturbing, as it reveals the erosion of due process when law is mediated through code. What is presented as "objective" calculation masks a normative decision: whose data is collected, how it is classified, and what thresholds are set.[33] Far from being neutral,

[27] Psycho-Pass (n 3).
[28] Lum and Isaac (n 6) 19.
[29] Kathrin Hartmann and Georg Wenzelburger, 'Uncertainty, Risk and the Use of Algorithms in Policy Decisions: A Case Study on Criminal Justice in the USA' (2021) 54 Policy Sciences 269.
[30] *Ibid* 269.
[31] Katia Schwerzmann, 'Abolish! Against the Use of Risk Assessment Algorithms at Sentencing in the US Criminal Justice System' (2021) 34(4) Philosophy & Technology 1883.
[32] *Ibid.*
[33] Ferguson (n 7).

these systems encode biases into legal decision-making under the guise of scientific precision.[34] Citizens are not recognised as autonomous agents capable of moral choice, but as data points subject to calculation and classification. To illustrate, if a particular demographic is disproportionately associated with crime in the data, the system may infer that this group is inherently more prone to offending. This creates a feedback loop of biased policing and sentencing, entrenching systemic discrimination and ultimately fuelling civil unrest.[35]

What is presented in *Psycho-Pass* as dystopian fiction is not as remote from our own legal debates as it first appears. The very tendencies dramatized by Sibyl, such as predictive policing, algorithmic risk assessments, and automated governance, are already being tested in real-world criminal justice systems.

Real World Parallels I: COMPAS and the Judicial Embrace of Algorithmic Risk

While the Sibyl System is fictional, its architecture closely resembles real predictive tools such as the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) algorithm, widely deployed across the United States to inform bail, parole, and sentencing decisions.[36]

In practice, the COMPAS algorithm assigns risk scores to criminal defendants based on factors such as criminal history, age, and socioeconomic background.[37] These risk scores are developed to assess recidivism risk at sentencing, but studies have revealed systemic flaws. COMPAS has misclassified Black defendants as "high risk" almost twice as often as white defendants,[38] while also exhibiting poor predictive accuracy as low as 20% in cases of violent crime.[39] Such systems entrench racial bias behind a veneer of objectivity, thereby concealing discriminatory practices within the authority of "data-driven" justice.[40]

The judicial reception of such tools is telling. In ***Loomis v Wisconsin***[41], the defendant challenged the use of COMPAS in sentencing on due process grounds, since the algorithm's

[34] Angwin and others (n 5).
[35] Lum and Isaac (n 6).
[36] Angwin and others (n 5).
[37] Angwin and others (n 5).
[38] Angwin and others (n 5). ProPublica's investigation revealed that COMPAS mislabeled Black defendants as high-risk nearly twice as often as white defendants who did not reoffend (45% compared to 23%).
[39] Angwin and others (n 5). COMPAS was accurate only about 61% of the time, and its accuracy dropped to just 20% when predicting violent crime.
[40] Ferguson (n 7).
[41] *State v Loomis* 881 NW2d 749 (Wis 2016).

inner workings were proprietary and unchallengeable.[42] In spite of these concerns, the court upheld its use, reasoning that COMPAS could provide useful information to promote accuracy and public safety, so long as judges were warned about its limitations and did not treat it as the sole basis for sentencing.[43] Yet this reasoning is deeply contradictory: by acknowledging the tool's opacity and risk of bias while still permitting its use, the court effectively tolerated a system in which defendants cannot meaningfully challenge the evidence against them, privileging efficiency over transparency and due process.

This reasoning reveals a jurisprudential shift: courts are increasingly willing to accept efficiency and predictive accuracy (however flawed) at the expense of transparency, accountability, and the defendant's right to challenge the evidence against them. If courts are willing to defer to flawed, opaque, and discriminatory algorithms, then transparency and accountability, core elements of the rule of law, risk collapsing. In such a system, judicial authority is no longer exercised through open legal reasoning but outsourced to proprietary systems controlled by private corporations. The danger is that law ceases to be the product of democratic institutions and becomes subordinated to the logics of technological vendors, thereby undermining both due process and the legitimacy of judicial decision-making.

This judicial willingness to embed opaque algorithms into sentencing practice parallels the logic of Sibyl: the substitution of legal reasoning with probabilistic forecasts, and the masking of normative decisions - who counts as "risky" - beneath the language of science. If courts accept black-box predictions as legitimate legal evidence, they risk eroding due process and hollowing out the very principles that distinguish justice from control.

Real World Parallels II: PredPol and the Feedback Loop of Policing

Another real-world parallel to Sibyl is PredPol,[44] a predictive policing software originally developed in partnership with the Los Angeles Police Department LAPD.[45] Similarly, PredPol uses historical crime data to forecast the time and location of future offences, supposedly enabling more efficient police deployment.[46] However, the effect of these efficient

[42] *Ibid* 757, 761. The methodology of COMPAS is kept a trade secret.
[43] *Ibid* 764-765.
[44] Lum and Isaac (n 6).
[45]Tim Lau, 'Predictive Policing Explained' (*Brennan Center for Justice*, 1 April 2020) <https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained> accessed 4 August 2025.
[46] *Ibid.*

deployments often reinforce existing patterns of racial profiling and over-policing.[47] As police patrol more heavily in neighbourhoods already marked as "high risk," more crimes are detected there, reinforcing the original prediction. As Lum and Isaac's study on Oakland demonstrated, such predictive systems reproduce and intensify existing racial disparities, creating what they call a "runaway feedback loop" of biased policing.[48] This feedback loop transforms biased historical data into a self-fulfilling prophecy, effectively criminalising entire communities rather than individuals.

The discriminatory consequences of such systems have been increasingly recognised.[49] Studies reveal that predictive policing disproportionately targets minority and low-income communities, magnifying pre-existing inequalities in policing.[50] Unlike traditional policing, where discretion is exercised by identifiable officers subject to oversight, predictive systems obscure responsibility by framing discriminatory outcomes as the inevitable result of "objective" data analysis.[51] The opacity of these systems undermines contestability, leaving those affected with little avenue to challenge the logic or accuracy of the predictions that shape their daily lives.

This dynamic closely mirrors Sibyl's logic. Just as Sibyl designates individuals as latent criminals based on statistical risk rather than actual conduct, PredPol treats neighbourhoods as inherently criminogenic on the basis of past data. Both systems collapse the distinction between probability and culpability: suspicion becomes perpetual, and risk is equated with guilt. By embedding bias into a predictive cycle, Sibyl and PredPol alike erode the foundational legal principle of the presumption of innocence, replacing it with a presumption of dangerousness.

[47] Lum and Isaac (n 6) 18. "We find that rather than correcting for the apparent biases in the police data, the model reinforces these biases".
[48] Lum and Isaac (n 6) 18. "Using PredPol in Oakland, black people would be targeted by predictive policing at roughly twice the rate of whites.Individuals classified as a race other than white or black would receive targeted policing at a rate 1.5 times that of whites. This is in contrast to the estimated pattern of drug use by race, where drug use is roughly equivalent across racial classifications".
[49] Ferguson (n 7) 43.
[50] Lum and Isaac (n 6) 18. "We find similar results (of targeted policing) when analysing the rate of targeted policing by income group, with low-income households experiencing targeted policing at disproportionately high rates. Thus, allowing a predictive policing algorithm to allocate police resources would result in the disproportionate policing of low-income communities and communities of colour."
[51] Danielle Citron and Frank Pasquale, 'The Scored Society: Due Process for Automated Predictions' [2014] 89 Washington Law Review 1.

Real World Parallels III: Surveillance Technologies and the Erosion of Privacy

Taken together, tools like COMPAS and PredPol reveal how predictive governance steadily shifts the law from adjudicating past conduct to managing future risk. Yet risk-based justice does not stop at policing or sentencing. Around the world, governments have begun experimenting with far more expansive systems of algorithmic surveillance and social control, extending predictive logic from criminal justice into the fabric of everyday life.

China: The Path Closest to Sibyl

China's Social Credit System[52] is perhaps the clearest real-world analogue to Sibyl. By aggregating financial records, social media activity, travel behaviour, and even interpersonal associations into a unified "trustworthiness" score,[53] the system pre-emptively sanctions individuals through travel bans, blacklists, and exclusion from opportunities.[54] Crucially, many of these sanctions are imposed automatically, without transparent procedures or avenues for contestation.[55] Like Sibyl's Crime Coefficient, these scores blur the line between prediction and punishment, collapsing due process into automated sanction. Legality is displaced by risk management, and individual dignity is subordinated to state-defined metrics of reliability.

United States: The Corporate-Sibyl

In the US, systems such as COMPAS[56] and PredPol[57] demonstrate how predictive governance can present structural bias under the guise of scientific objectivity.[58] These tools are developed and controlled by private corporations, shielded from scrutiny by trade secrecy.[59] Courts, as in ***Loomis v Wisconsin***[60], have shown willingness to defer to these opaque systems, even when defendants cannot challenge their inner workings.[61]

---

[52] Rogier Creemers, 'China's Social Credit System: An Evolving Practice of Control' (*SSRN*, 9 May 2018) <https://ssrn.com/abstract=3175792> accessed 4 August 2025.
[53] Xin Dai, 'Toward a Reputation State: The Social Credit System Project of China' (*SSRN*, 24 Jun 2018) <https://ssrn.com/abstract=3193577> accessed 4 August 2025.
[54] *Ibid.*
[55] Creemers (n 53) 12–14.
[56] Angwin and others (n 5).
[57] Lum and Isaac (n 6).
[58] Ferguson (n 7).
[59] *Loomis* (n 43).
[60] *Loomis* (n 42).
[61] *Loomis* (n 43).

However, the court still remains cautious as to the use of COMPAS.[62] Unlike China's state-driven Social Credit, this is a privatised Sibyl, where proprietary algorithms, rather than state institutions, increasingly mediate judicial authority.

European Union: Guardrails Against Sibyl

The European Union charts a different path, emphasising safeguards rather than uncritical adoption. The **General Data Protection Regulation (GDPR)**[63] enshrines the right not to be subject to purely automated decision-making **(Article 22)**.[64] The EU also formally regulates AI under the **Artificial Intelligence Act**[65], designating predictive policing and social scoring as "high-risk" systems subject to stringent compliance requirements.[66] Rather than embracing prediction as inevitable, the EU foregrounds contestability, transparency, and proportionality. This represents a deliberate attempt to preserve Fuller's "inner morality of law"[67] by ensuring that law remains open, contestable, and rooted in human judgment.

This stands in stark contrast to Sibyl's unchecked dominance, demonstrating that law need not passively surrender to technological determinism. Instead, it can actively constrain the terms on which predictive technologies enter public governance.

Three Paths, One Question

These trajectories: China's statist model, America's corporate model, and Europe's regulatory model, reveal divergent futures of algorithmic governance. Yet each raises the same fundamental question dramatized in *Psycho-Pass*: can legality survive when judgment is replaced with classification? Whether through state control, corporate opacity, or regulatory safeguards, the struggle is not about technology alone but about law's ability to retain legitimacy in an age where code increasingly governs conduct.

[62] *Loomis* (n 44).
[63] Regulation (EU) 2016/679 (General Data Protection Regulation) [2016] OJ L119/1.
[64] Regulation (EU) 2016/679 (General Data Protection Regulation) [2016] OJ L119/1, art 22.
[65] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) [2024] OJ L 1689.
[66] Regulation (EU) 2024/1689 (Artificial Intelligence Act) [2024] OJ L 1689, Annex III.
[67] Fuller (n 26).

What unites these systems is the substitution of legal judgment with algorithmic classification. Whether in the form of Sibyl's Crime Coefficient, COMPAS's recidivism scores, PredPol's crime maps, or China's trustworthiness ratings, law risks being redefined as a mechanism of behavioural management rather than adjudication of rights and wrongs. In Fuller's terms, the "inner morality of law"[68] is jeopardised: norms lose their stability and transparency when they are supplanted by shifting algorithmic thresholds. Predictive systems present themselves as neutral tools, yet they embed normative assumptions about what constitutes "trustworthiness," "risk," or "criminality." In doing so, they collapse justice into control, and legality into code.

Conclusion: Toward Human-Centered Legal Futures

When justice becomes data, law becomes control. That is Sibyl's dystopia, and our warning.

Psycho-Pass warns that a legal system governed by prediction, not principle, may sacrifice justice for predictability. As predictive tools proliferate, society must guard against systems that replace deliberation with data and accountability with automation.

To preserve the Rule of Law, we must insist on:

1. Transparent and contestable algorithms
2. Human judgment in decision-making
3. Protection of due process and moral agency

Ultimately, justice requires more than accuracy. It requires empathy, fairness, and the recognition that law must serve humanity as a mere tool, not the other way around.

[68] Fuller (n 26).