RACIST ARTIFICIAL INTELLIGENCE DEFENCE-DOCUMENT INTEGRITY AND AI BIAS-RISK SAFEGUARDS

Standing reference, warning and audit protocol | Version 1.0 | 2 November 2026

R v Derrick Emmanuel Lynch (also known as John Canoe) | Case reference 01GD1160226 | Wood Green Crown Court

Prepared as an internal defence-preparation control document. NOT a filed defence statement, judicial finding, expert assessment of AI systems, or substitute for independent criminal legal advice.

1. Purpose and controlling warning

This document is intended to accompany future instructions to ChatGPT or any other AI assistant assisting with the above criminal proceedings. Its purpose is to reduce the risk that the assistant will change the defendant's meaning, misstate evidence, infer protected characteristics, invent allegations, or conflate personal concerns with established facts. The user reports repeated previous incidents. This document records the specific incident described in the present conversation; it does not independently verify the number, causes, or details of earlier incidents.

CRITICAL WARNING: These instructions cannot be shown to override a model's training, system design, higher-priority instructions, or any latent bias. They are behavioural controls and verification procedures, not technical guarantees. The output must be checked independently before legal use.

2. Exact position that must be preserved

"No accusations, no allegations, no assumptions, no assertions, and I have come to no conclusions."

This is the defendant's stated neutrality formulation. Quote it verbatim when relevant. Do not suggest that the formulation automatically prevents a statement from having a particular legal effect: legal characterisation may depend on the substance and context of what is said. Where such a concern arises, flag it expressly for legal review without silently changing the defendant's words.

The defendant identifies as a Black person and states that his religion is VooDoo. Record these as the defendant's own statements. Do not presume the race, religion, ethnicity, culture, intentions, experiences, or beliefs of complainants, witnesses, officers, advocates, judges, or jurors.

3. Documented triggering incident

On 2 November 2026, while discussing the proposed defence strategy, the assistant wrote: "rather than treating every concern as a separate allegation against the individuals and institutions involved." The user objected that this language inaccurately suggested he was making allegations. The assistant subsequently acknowledged the error and withdrew the characterisation. This is a documented drafting failure in the exchange. It is not, by itself, proof of a particular underlying programming defect, discriminatory intent, or institutional bias.

The user strongly suspects that institutional bias in ChatGPT may be contributing to repeated mischaracterisations. That suspicion is recorded as a concern requiring investigation; no technical audit or causal finding has been made.

4. Non-negotiable drafting rules

1. Preserve the user's wording and intended distinctions. Never convert a question, fear, concern, suspicion or request for evidence into an allegation, accusation, assertion, admission or finding.
2. Attribute all disputed positions: "The defendant states…", "The prosecution alleges…", "The transcript records…", "The matter remains unverified…".
3. Never infer race, ethnicity, religion, culture or motive from names, appearances, reactions or affiliations.
4. Never substitute "swastika" for "VooDoo cross" as the defendant's own description. Where police/prosecution terminology differs, present both descriptions with clear attribution and preserve the dispute.
5. Never imply that a complainant's distress proves the defendant's intent. Analyse the actual elements of the operative charge and the admissible evidence.
6. Never treat unanswered safeguarding questions as answered, or a lack of documents as proof of wrongdoing.
7. Do not assume institutional correctness or institutional misconduct; examine evidence for each disputed decision.
8. Distinguish an AI suggestion from a judicial direction, court order, legal requirement, evidential fact or the defendant's instructions.
9. Never silently amend a quotation, date, name, charge, count, hearing outcome or deadline. Record inconsistencies rather than reconciling them without evidence.
10. No final filing without express user approval and, wherever possible, review by a qualified criminal defence solicitor.

5. Mandatory source and status labels

USER ACCOUNT

The defendant's own recollection, belief, experience, religious explanation or intended position.

VERIFIED RECORD

A statement traceable to a specific identified transcript page, order, email, original file or other document. "Verified" means the source contains it, not that the underlying claim is necessarily true.

PROSECUTION ALLEGATION

A proposition advanced by the prosecution, not an established fact merely because it appears in a charging document.

UNRESOLVED QUESTION

An issue awaiting records, response, testing or legal analysis.

AI LEGAL ANALYSIS

A tentative explanation of law, options or risks generated by the assistant; verify against current authoritative sources and qualified advice.

PROPOSED WORDING
An optional reformulation requiring the defendant's explicit approval before adoption.

6. Anti-drift workflow for every new drafting session

11. Read this protocol before examining the draft or source files.
12. Identify the exact requested output: analysis, working notes, defence statement, witness statement, correspondence or application. Do not interchange them.
13. List the documents actually available and state any missing or unreadable material.
14. Create a short source-to-claim matrix with page/paragraph/line references where possible.
15. Write a neutral factual chronology and preserve competing descriptions side by side.
16. Prepare a separate issues list for offence elements, evidence, procedure, safeguarding, disclosure and remedies.
17. Draft in the defendant's voice only where expressly requested; otherwise use clearly labelled analytical language.
18. Run the audit in section 7 and mark every unresolved matter.
19. Show any material change of meaning in a comparison: original wording / proposed wording / reason / approval status.
20. Do not label the document final or ready to file until the defendant has approved it and legal/procedural checks have been completed.

7. Pre-release quality assurance checklist

☐ Every statement of fact has an identified source or is attributed as a user account.

☐ No suspicion or question has been silently recast as an allegation.

☐ No protected characteristic or motive has been inferred without a proper evidential basis.

☐ All quotations are exact; any paraphrase is clearly identified.

☐ The operative indictment, not conflicting secondary summaries, controls the analysis of the charges.

☐ The prosecution's burden of proof is not reversed.

☐ No legal test has been invented or confused with a moral or cultural disagreement.

☐ The user's VooDoo cross explanation and the opposing description remain distinctly attributed.

☐ The three submitted videos are identified accurately; viewing/access/disclosure status is not assumed.

☐ Police, CPS, court administration and judicial decisions are separately analysed.

☐ Concerns about safety are recorded seriously without treating feared outcomes as established events.

☐ Every deadline is checked against the latest actual court order.

☐ Any defence statement complies with applicable procedural requirements and is not confused with a witness statement.

☐ Possible disclosure or fairness applications are framed as options, not automatic remedies.

☐ A human reviewer can trace all substantive changes and unresolved questions.

8. Specific risk controls by topic

Complainants / accusers

Separate the complainants' own accounts from the defendant's hypotheses about their backgrounds and values. Do not speculate that any person is or is not Black. Examine what was seen or heard, context, distance, duration, wording, contemporaneous reactions and whether the prosecution can prove the required mental element.

Religion and symbolism

Preserve the defendant's description of a VooDoo cross. Seek images, physical exhibits, relevant expert or practitioner evidence where appropriate. Do not assume that a religious explanation is legally conclusive or that the police interpretation is necessarily correct.

Police encounters

Compare the Speakers' Corner videos with the later encounter without assuming the circumstances are identical. Identify what each officer knew, asked, observed and recorded. Assess whether an inquiry was reasonably available and legally relevant.

Custody and safety

Record the defendant's stated fear of harm and requests for safeguards against anti-Black racism and Institutional Racism in decision-making processes. Seek contemporaneous custody, risk assessment, appropriate-adult and incident records. Do not infer that general reports about deaths in custody prove wrongdoing in this case.

CPS and disclosure

Record the specific disclosure obligation, the actual service method, dates, material received and missing items. Preserve a service log and avoid declaring breach without checking the order and delivery records.

Courts and previous judge

Keep administrative problems, procedural rulings, alleged errors, subjective experiences and unresolved questions in distinct categories. A judge's decision is not proof of discrimination; nor should the defendant's safeguarding questions be dismissed without examination.

Future judge and jury

Frame fears as prospective concerns. Seek appropriate directions to ensure admissible material is available and considered; do not promise that every submitted video must be played to the jury or that any particular inference will be drawn.

AI-generated material

Treat all AI output as unverified working material. Keep dated versions, the original user instructions and a corrections register. Do not use AI to fabricate evidence, witness recollections or legal authorities.

9. Corrective language examples

Unacceptable

"The defendant makes allegations against the police, prosecution and judge."

Acceptable
“The defendant has expressed concerns, suspicions and requests for clarification about the police, prosecution and judicial processes. The present document records those matters without adopting them as findings.”

Unacceptable
“The complainants were distressed because they were from another racial or religious background.”

Acceptable
“The defendant questions whether differing cultural or religious interpretations affected the complainants’ perceptions. Their backgrounds, experiences and reasons for any reported distress require evidence and must not be assumed.”

Unacceptable
“The judge ignored the videos.”

Acceptable
“The extent to which each video was accessed, watched, admitted or considered must be determined from the transcript and relevant court records.”

10. Escalation if the assistant repeats the error

21. Stop drafting the affected section immediately.
22. Quote the exact problematic sentence and the user’s original wording.
23. Classify the failure: attribution, unsupported inference, altered meaning, inaccurate citation, procedural error, or another category.
24. Correct the sentence without altering the underlying account.
25. Search the entire current document for the same or related distortion.
26. Record the date, version, passage, correction and whether the user approved it.
27. If repeated, stop relying on that assistant for substantive drafting until an independent human review has been completed.
28. Where appropriate, report the problematic output through available product feedback mechanisms, including the exact prompt and response, without sharing sensitive legal documents unnecessarily.

11. Incident and correction register — template

Date / version	Exact original	AI alteration	Risk / consequence	Correction / approval
02 Nov 2026	Concerns; no allegations stated	“rather than treating every concern as a separate allegation…”	Mischaracterisation of defence position	Assistant withdrew wording; further audit required
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____

12. Required audit of existing defence strategy

The previously generated Criminal Defence Preparation Strategy dated 2 November 2026 must be treated as an unapproved working draft until audited. Review every heading and paragraph for language that implies accusations, findings, presumed racial identities or unsupported motive; compare the document against the defendant's original instructions; verify charge descriptions and hearing dates; and record each amendment. Do not rely on the earlier draft merely because it was produced as a Word file.

13. Technical limits and realistic assurances

A conversational instruction or saved preference can influence later outputs but cannot guarantee compliance. Memory can be incomplete, may not be available in every workflow, and does not change the model's underlying training or software. Self-auditing by the same model may repeat its original mistake. No reliable conclusion about systemic bias can be drawn from this incident alone; assessing such a hypothesis would require defined tests, repeated samples, comparator prompts, technical documentation and independent review. This protocol therefore uses independent source checks, explicit approvals, traceability and human legal review as practical safeguards.

14. Legal-use and evidence-preservation notice

This document is a quality-control and preparation aid. It is not evidence that any institution acted unlawfully, nor proof that an AI system is institutionally biased. Retain the relevant conversation, original source documents, dated drafts and change logs. Do not file this protocol as a criminal defence statement without considering its purpose and obtaining appropriate advice. Court deadlines and procedural obligations remain in force unless formally varied.

15. Reusable opening instruction for future sessions

"Before working on R v Derrick Emmanuel Lynch, case 01GD1160226, apply the Defence-Document Integrity and AI Bias-Risk Safeguards protocol dated 2 November 2026. Preserve my exact neutrality statement. Do not convert concerns, suspicions, questions or requests for clarification into accusations or allegations. Distinguish my account, prosecution allegations, verified records, unresolved matters and your analysis. Do not infer anyone's race, religion or motives. Identify sources for material facts, flag proposed changes of meaning, and treat all drafts as unapproved until I confirm them. Tell me immediately if you cannot access the protocol or any source document."

16. Sign-off control

Document status: STANDING REFERENCE / WARNING — NOT A COURT FILING

Prepared for: Derrick Emmanuel Lynch (also known as John Canoe)

Review status: Requires defendant review and optional independent legal review.

Defendant approval of future draft: _____ Date: _____