

Jerry Falade

The Ghost of Phillis Wheatley: AI Detection and the Crisis of Black Authorship

ABSTRACT

In 1772, the Black poet Phillis Wheatley had to be examined by eighteen prominent white men in Boston, including Governor Thomas Hutchinson, to determine whether an enslaved African girl could have genuinely written the poems attributed to her. Almost three centuries later, the tribunal has better technology and AI detection software is being used to ask successful Black writers the same question: *Hey, you are Black. Did you really write this book yourself or did ChatGPT write it for you?* This paper traces a genealogy of contested authorship from Wheatley to the age of artificial intelligence, arguing that the contemporary reliance on algorithmic verification reproduces historical structures of intellectual gatekeeping. The machine makes the accusation. The Black author has to prove their humanity.

This paper examines how racialized and linguistically non-normative authors are systematically mandated to authenticate their intellectual labor—often provoked by dominant white literary gatekeepers who subject Black authors to scrutiny rarely applied to their white contemporaries. For example, in Falade’s case, a white journalist, who is also a novelist, obtained his unpublished manuscript without his knowledge or consent, ran it through AI detection software, emailed the results to his creative partners, and published the resulting assessment in an article that was subsequently edited.

Recent research has demonstrated that AI detectors can disproportionately misclassify writing by non-native English speakers (Liang et al., 2023). I’ll use this racially-misaligned technological context to examine five literary controversies that emerged in 2026: Jerry Falade’s *Call Me, I’ll Hide the Body*, H.M. Wolfe’s *Daggermouth*, Jamir Nazir’s Commonwealth Prize-winning “The Serpent in the Grove,”

Mia Ballard's *Shy Girl*, and Haitian-Canadian novelist Thélyson Orélien's *C'était ça ou mourir*. Five names, all People of Color. The consequences of these AI detectors weaponized by "White cultural supremacist gatekeepers" have ranged from cancelled publication and lost multimillion-dollar deals to removal from major literary-prize consideration and public accusations amplified by White gatekeepers and the PR teams of certain AI detection software who seem to enjoy a fairly good amount of publicity whenever a Black author is unfairly accused of writing with AI.

Using Falade's 2026 essay "Me and My AI Against the World" as a central contemporary text, I will develop "technogatekeeping" to describe the use of technological systems as instruments of cultural and institutional gatekeeping: systems through which access, legitimacy, authorship, and credibility are increasingly adjudicated by algorithmic classifications rather than documentary evidence alone. In the context of Black and diasporic authorship, technogatekeeping occurs when algorithmic systems become intermediaries between the writer and the institutions empowered to recognize that writer as legitimate. I argue that this produces linguistic erosion as non-native English writers increasingly face pressure to modify human stylistic differences to appear sufficiently "human" to profit-oriented AI detectors. This paper asks what changes when Black authorship becomes something to always be contested and proven, and why the Western gatekeepers put the burden of proof on marginalized writers instead of the accusers.

Phillis Wheatley had to prove that a Black girl could write in the 18th century. In 2026, Black writers are again being asked to prove what they wrote. This time around, the gatekeepers are not eighteen white men, rather a flawed AI detector that seems to gain more popularity by bringing down Black writers.

New century. Different Tribunal. Familiar question: Did your inferior Black brain write this book or did you get external help?