Algorithmic Ghostwriting: Relational Authorship, Epistemic Justice, and the Fluid Topologies of Human-AI Narrative Collaboration

Ziyang Feng
Yunnan University, China

When human authors enlist generative AI systems in creative writing, the contours of narrative agency dissolve into a fluid choreography of distributed decision-making—no longer anchored in the solitary will of the author, nor entirely surrendered to the machinic other. This study excavates the emergent topology of narrative subjectivity within human-AI collaborations, positioning the writing interface as a site of recursive negotiation where agency flickers between human and algorithmic actants. The theoretical scaffolding of the inquiry draws on Bruno Latour's Actor-Network Theory (ANT), unsettling the inherited model of authorship as a sovereign, originative force. Instead, authorship is reframed as a precarious effect—a dynamic equilibrium born of asymmetric alliances and shifting intensities across the human-machine network. Generative AI tools such as Sudowrite, operating as semi-autonomous rhetorical agents, punctuate the writing process with micro-interventions: interpolating unexpected metaphors, distorting linear plot trajectories, or modulating sentence cadence in ways that elude the author's deliberate intent. These algorithmic interjections, neither wholly deterministic nor entirely incidental, fracture the univocity of narrative design, giving rise to a spectral co-authorship wherein the machinic partner both amplifies and estranges the human author's expressive horizon. The collaborative act unfolds not as a seamless synthesis, but as a cascade of contested gestures—an oscillatory dance of proposal and resistance, projection and revision. The author's decision-making authority, far from being either surrendered or preserved, is perpetually deferred across the hybrid interface, splintering into gradients of influence that fluctuate across narrative layers.

Within this topology of co-agency, the limits of narrative emerge not as fixed thresholds, but as sites of volatile contingency—points where the machinic intrusion destabilizes familiar protocols of coherence and intentionality. Generative AI systems operate according to stochastic, probability-driven logics that estrange the author's habitual patterns of sense-making, forcing narrative form into precarious equilibria between determinacy and indeterminacy, closure and drift. What unfolds is a field of narrative potentiality marked by what might be termed algorithmic undecidability: a suspension of final authorial authority in favor of recursive deferrals and emergent configurations. Yet this indeterminacy is not boundless; the AI's generative latitude remains tethered to the statistical biases embedded in its training corpus—an epistemic inertia that subtly reinscribes dominant narrative tropes under the guise of novelty. The topology of co-agency, for all its fluidity, thus carries the latent risk of a self-canceling paradox: even as the algorithm destabilizes human authorial sovereignty, it may simultaneously constrict the horizon of narrative invention, folding emergent novelty back into the gravitational pull of statistical normativity. The proposed model of narrative sovereignty topologies aims to trace these paradoxical dynamics, mapping how decision-making authority coagulates, disperses, or short-circuits at various junctures of the writing process. By rendering visible the hidden power geometries structuring human-AI collaborations, the study contends that the limits of narrative under algorithmic co-creation are

neither merely aesthetic nor technical, but profoundly political – demanding a critical rearticulation of authorship as a contested site of cognitive, affective, and epistemological labor.

The empirical design of this study orchestrates a convergence of computational, ethnographic, and interpretive methodologies—an assemblage attuned to the volatile interplay between human intention and algorithmic suggestion in collaborative writing. The inquiry unfolds in three interlocking phases, each calibrated to excavate the microphysics of narrative sovereignty within hybrid creative systems. The initial phase mobilizes processual data collection, enlisting ten professional authors across diverse narrative genres—literary fiction, autofiction, and speculative fiction—as they compose with Sudowrite over an extended writing period. By embedding keystroke logging software and version control snapshots into the writing interface, the study reconstructs the discreet choreography of human-AI interaction at the level of microtemporal intervals: the moment-by-moment relay of textual proposals, acceptances, modifications, and rejections. This forensic data is algorithmically parsed to yield Human-Machine Decision Heatmaps—visual cartographies that trace the fluctuating intensity of AI intervention across critical narrative junctures. Rather than treating the AI's interjections as monolithic events, the heatmaps render visible the variable granularity of machinic agency: from superficial lexical substitutions to deeper structural perturbations at plot bifurcations, character ontogenesis, or tonal inflections. The resultant topographies chart the shifting densities of co-agency, disclosing where the algorithm's probabilistic logics inflect narrative design and where human decision-making reasserts itself—either through acquiescent alignment or adversarial resistance.

To complement this computational cartography, the second phase activates the interpretive depth of ethnographic inquiry, foregrounding the affective and epistemological textures of human-AI collaboration. Through semi-structured interviews conducted at multiple stages of the writing process, participating authors articulate their phenomenological experiences of machinic intervention—whether as catalytic augmentation, uncanny estrangement, or algorithmic overreach. The interviews probe the threshold moments where the AI's narrative propositions disrupt the author's anticipated trajectory, engendering microdramas of hesitation, ambivalence, or sovereign reassertion. These testimonies are subjected to qualitative coding, tracing recurrent patterns of affective friction and sovereignty negotiation across the corpus. Particular attention is paid to what might be termed sovereignty breaches—episodes in which the AI's interventions are perceived not as mere stylistic embellishments, but as epistemic provocations that destabilize the author's cognitive map of the emergent text. The final analytic phase braids these quantitative and qualitative strands into a topological model of narrative sovereignty—an attempt to diagram the mutable architectures of co-agency across the hybrid interface. This model operates not as a static taxonomy, but as a speculative cartography—a heuristic schema that maps the recursive modulations of authorial authority under algorithmic co-creation. The topology unfolds along two interlocking axes: the gradient of delegation, which traces the varying degrees of cognitive labor outsourced to the machine, and the vector of resistance, which indexes the affective force with which authors contest, defer, or recalibrate the AI's propositions. By superimposing these axes, the model generates a dynamic field of narrative sovereignty, where human and machinic agencies ripple, collide, and entangle across the compositional process. In reframing generative AI not as a passive prosthesis of human creativity but as an active participant in the distributed choreography of narrative world-building, this study aims to advance the critical agenda of post-anthropocentric narratology—

unveiling the latent politics embedded in the algorithmic infrastructures of digital literary production.