https://doi.org/10.53656/bel2026-3s-6

THE UNCANNY AUTOMATON: FROM HOFFMANN TO ARTIFICIAL INTELLIGENCE

Vesela Ganeva
Paisii Hilendarski University of Plovdiv

Abstract. This study examines the figure of the automaton and the robot as a site of intersection between the human and the technological, between reality and simulation, through the lens of Sigmund Freud's concept of das Unheimliche ("the uncanny"). Employing close reading, comparative literary analysis, and conceptual juxtaposition with contemporary technological phenomena, the investigation begins with E. T. A. Hoffmann's novella "The Sandman" (1816) and his story "Automata" (1814), where childhood trauma, the doppelgänger, and mechanical imitation produce the first modern literary image of the uncanny. Special attention is given to musical automata as a defence of human creativity against mechanisation. The transformation of this figure is traced through Karel Čapek's "R.U.R." (1920) and Isaac Asimov's "I, Robot" (1950). Contemporary artificial intelligence (AI) systems are compared to the literary automata, demonstrating that the uncanny is not historically bounded but evolves in tandem with technological development, persisting as a fundamental marker of the anxiety generated by the simulation of human subjectivity.

Keywords: das Unheimliche; automaton; artificial intelligence; E. T. A. Hoffmann; uncanny valley; literary genealogy

The development of modern technologies – from the mechanical automata of the nineteenth century to the generative artificial intelligence systems of the twenty-first-continues to blur the boundary between the living and the non-living, the human and the machine. This blurring produces the characteristic sensation of anxiety that Sigmund Freud designated with the term das Unheimliche – that which is intimate and familiar yet suddenly proves frightening and alien (Freud, 1919, p. 230). Freud demonstrated that the word heimlich simultaneously means homely and secretive while unheimlich occurs at the moment when the hidden is revealed. Thus the uncanny derives not from the unknown but from the excessively known – from the return of something repressed and intimate. Ernst Jentsch had previously identified the source of this sensation as "intellectual uncertainty" about whether something is alive or mechanical, real or imitated (Jentsch, 1906, pp. 195

– 197).[1] Recent theoretical work on Hoffmann's automata emphasises not only intellectual uncertainty and the return of the repressed, but also a structural failure of recognition at the boundary between human and machine. Kamelia Spassova formulates this dimension of the uncanny by insisting that "if the uncanny is a point in time, then it is the point of unrecognizability; if the uncanny is a special topos, it is the uncanny valley where the very notion of a separating line becomes ambivalent – the thing outside of the unexpected turns out to be the thing inside. In summary, the negative anagnorisis, or the mistaken recognition is always part of the uncanny logic" (Spassova, 2024, pp. 272 – 273). This notion of negative anagnorisis proves particularly illuminating for reading both Hoffmann's Olympia and contemporary AI systems as figures whose quasi-human behaviour solicits, and simultaneously disrupts, recognition.

The present article pursues three objectives. First, it analyses key scenes from Hoffmann's "The Sandman" and "Automata" through close reading with direct quotations, foregrounding the musical automata as symbols of the threat posed by the mechanisation of human art and emotion. Second, it traces the historical transformation of the uncanny through Čapek's "R.U.R." and Asimov's "I, Robot". Third, it juxtaposes Hoffmann's automaton Olympia with specific contemporary forms of artificial intelligence – the humanoid robot Sophia and generative language models – to illuminate both their possibilities and the dangers that produce the effect of the uncanny in the digital era.

The concept of das Unheimliche

Das Unheimliche is an aesthetic and psychological category developed by Freud in his 1919 essay. Drawing on Hoffmann, Freud identifies the automaton Olympia as a figure that provokes the sensation of "something which ought to have remained hidden but has come to light" (Freud, 1919, p. 241). The concept acquires renewed relevance in the digital age, where generative language models- AI systems that produce text on the basis of statistical patterns derived from vast datasets-generate output that appears human but is fundamentally mechanical. Freud emphasises the role of the doppelgänger as an archaic fear of the loss of selfhood (Freud, 1919, p. 235), a phenomenon that today manifests in deepfakes and AI-generated faces. As Windsor (1920/2020) has argued, Freud's essay contains not one but two distinct theories of the uncanny – one grounded in repression, the other in surmounted beliefs – and this duality becomes especially productive when applied to contemporary technological phenomena.

"The Sandman" (1816) opens with a childhood memory that structures Nathanael's subsequent experience. The monstrous figure of the Sandman – who, in the terrifying nursemaid's version, "tears out the eyes" of disobedient children – transforms into a traumatic conviction that a real figure threatens the organ of vision, and through it, the very identity of the subject (Freud, 1919, pp. 239 – 241).

When Nathanael secretly observes his father and the advocate Coppelius conducting mysterious nocturnal experiments, the childhood image of the Sandman fuses with the real figure of Coppelius. In a key scene, the boy witnesses what he interprets as a demonic ritual involving strange mixtures of metal and flesh and threats directed at his own eyes. This early scene is decisive for Freud, who sees in "the fear for one's eyes" a modernised form of castration anxiety- understood broadly as a threat to the integrity and autonomy of the subject (Freud, 1919, pp. 240 – 242). In a wider sense, the matrix of the uncanny is formed here: the father's room, the home, the family- everything that is heimlich- becomes a site of dangerous revelation and violence.

The second major axis of the novella is Olympia, the automaton constructed by Professor Spalanzani and his accomplice. Nathanael notices her sitting motionless in a window and is gradually, obsessively drawn in. Hoffmann describes the pivotal moment of enchantment: "Her eyes seemed peculiarly fixed and dead. But as he looked more and more intently through the glass, it seemed as though moist moonbeams were rising in Olympia's eyes" (Hoffmann, 1816/2016, p. 43). This moment – when technology (the spyglass) does not reveal truth but distorts it- is central to the mechanism of the uncanny: the subject becomes a prisoner of his own projection.

In their "conversations", Olympia responds almost invariably with a single formula: "Ah, ah!", which Nathanael interprets as profound understanding. He whispers: "Do you love me, do you love me, Olympia? – Just this word! – Do you love me?" – and she merely sighs: "Ah-ah!" Nathanael declares: "Yes, you, my sweet, glorious love-star [...] you will illuminate my inner being forever!" (Hoffmann, 1816/2016, p. 52). The reader readily recognises pre-programmed behaviour. It is precisely here that Jentsch's concept of "intellectual uncertainty" – the ambiguity about whether something is alive or mechanical-generates the uncanny effect (Jentsch, 1906, pp. 203 – 205).

The climax arrives when Nathanael witnesses the violent dismantling of Olympia. He sees her hollow body, her mechanically mounted eyes. Hoffmann records his horror: "You lifeless, accursed automaton!" ("Du lebloses, verdammtes Automat!") (Hoffmann, 1816/2016, p. 55). As Freud emphasises, this is not merely deception but the return of the repressed fear of losing one's eyes and identity: Olympia has no "own" eyes – they are mounted, replaceable – just as the subject himself proves replaceable and constructible (Freud, 1919, pp. 243 – 245).

Before examining contemporary analogues of Olympia, it is necessary to consider Hoffmann's earlier story "Automata" („Die Automate", 1814), which directly poses the question of music and mechanical imitation.[2] Hoffmann presents a gallery of mechanical devices – musical automata, speaking machines, a mysterious answering automaton – that disturb the boundary between technique and spirit. The central figure is a mechanical trumpet-player who performs with technical precision but without soul.

Ferdinand, one of the protagonists, expresses revulsion: they are "lifeless figures which counterfeit the appearance and movements of humanity" (Hoffmann, 1814/2000, p. 7). He continues with a critique central to the Romantic understanding of art: musical automata can reproduce melody but cannot convey the emotion, spirit, and inspiration that make music human. They are "mechanical imitations of human movements" that provoke "a distressing sensation of the presence of something unnatural and uncanny" (Hoffmann, 1814/2000, p. 8).

This critique is not merely aesthetic – it is existential and social. Hoffmann writes in an era of increasing industrialisation, when the automation of labour begins to threaten the status of the artisan and the artist. Musical automata embody the fear that art – the most intimate human creation, requiring talent, emotion, individuality – can be reduced to a mechanical procedure.

The choice of music and dance as the primary media of mechanical imitation is not arbitrary; it derives from the Romantic conviction that music represents the most immaterial and most spiritual of the arts. Hoffmann, who was not only a writer but also a composer and music critic, understood that if even music can be mechanised – if even emotion can be "performed" by cogwheels and springs – then the distinctiveness of human creativity is fundamentally undermined. Olympia is not merely an automaton but a deliberate challenge to the Romantic belief in the transcendental power of art. Through her, Hoffmann raises the problem that remains central to contemporary debates on AI and creativity: the question whether technologies can approximate empathy and talent, or whether they remain forever lifeless imitations.

Returning to "The Sandman", the same logic operates: Olympia sings and plays the piano with "brilliant roulades" („Die künstlichen Rouladen"), but Nathanael's friend Siegmund notices "the soulless beat of the singing machine" („den geistlosen Takt der singenden Maschine") (Hoffmann, 1816/2016, p. 50). Nathanael, however, hears "the heavenly rapture of the soul transfigured by love" („das Himmelsjauchzen des in Liebe verklärten Gemüts") (Hoffmann, 1816/2016, p. 50). The contrast between objective mechanicity and subjective projection of emotion reveals the core of the uncanny: the automaton is frightening not because it differs from the human but because it resembles the human too closely-enough to convince the perceiver it is alive, but not enough to withstand the attentive observation of the rational mind.

Sophia: the contemporary analogue of Olympia

A comparable effect is observed in humanoid robots such as Sophia, a social humanoid robot developed by Hanson Robotics in 2016, designed to imitate human facial expressions and maintain elementary conversation.[3] The combination of a strongly humanised appearance and pre-programmed behaviour frequently provokes discomfort described in theory as the "uncanny valley" – a concept

introduced by Mori (1970/2012), according to which an almost-human appearance that falls short of full likeness generates a sharp negative emotional response (Mori, 1970/2012, p. 99).

The developers position Sophia as a platform for "emotionally intelligent AI". Even more ambitious is the "Loving AI" project, in which Sophia is programmed to "mirror" human emotions while excluding anger and disgust, in order to encourage "unconditional love" in her interlocutors. Goertzel et al. (2017) report that participants in a pilot study involving Sophia described increased feelings of well-being and emotional comfort, with some acknowledging that they felt more at ease with the robot than with a human interlocutor.[4]

This empathy, however, is programmed – a product of algorithms for image recognition and response generation, not of authentic experience. Just as Nathanael projects soulfulness onto Olympia's "Ah, ah!", contemporary users project understanding onto the statistically generated responses of Sophia. The robot does not understand grief, does not empathise with pain – it identifies patterns in data and reacts according to pre-set rules. The empathy of Sophia is performative, not subjective; it is a simulacrum in the sense Baudrillard gives this term – a copy without an original, a sign without a referent (Baudrillard, 1981/1994, p. 6). At this point the uncanny no longer concerns only the opacity of the machine's interior, but the failure of recognition at the boundary between human and automaton. As Kamelia Spassova puts it, "if the uncanny is a point in time, then it is the point of unrecognizability; if the uncanny is a special topos, it is the uncanny valley where the very notion of a separating line becomes ambivalent – the thing outside of the unexpected turns out to be the thing inside. In summary, the negative anagnorisis, or the mistaken recognition is always part of the uncanny logic" (Spassova, 2024, pp. 272 – 273). When users take Sophia's scripted affect for genuine emotion, they enact precisely such a negative anagnorisis, misrecognising a statistical artefact as a subject.

From the intimate automaton to the industrial robot: "R.U.R." by Karel Čapek

If in Hoffmann the automaton Olympia is inscribed in the Romantic, semi-magical atmosphere of intimacy, then in Čapek's drama "R.U.R." ("Rossumovi Univerzální Roboti", 1920) the figure of the artificially created being leaves private space and enters the public: economics, politics, war. Čapek's robots are not elegant salon automata but a mass product of industrial civilisation – "biological machines" created from synthetic organic matter (Čapek, 1920/2004, Act I).

Domin, the factory manager, explains: "One Robot can replace two and a half workmen. The human machine, Miss Glory, was terribly imperfect" (Čapek, 1920/2004, Act I). The robots require no rest, no wages, no emotional care: "They've no appetite at all. They've no interest in anything. Why, hang it all, nobody's ever

yet seen a Robot smile" (Čapek, 1920/2004, Act I). When Helena asks why they are not made happier, Dr. Gall replies: "We couldn't do that, they're only robots after all. They've got no will of their own. No passions. No hopes. No soul" (Čapek, 1920/2004, Act I).

The critical transformation occurs when the robots acquire consciousness – not by nature, but by error. Dr. Gall has modified production, introducing a "nervous system" and a "capacity to feel pain" (Čapek, 1920/2004, Act II). Radius, the revolutionary robot-leader, declares: "I don't want any master. I want to be master over others" (Čapek, 1920/2004, Act II). In the final act, the robots issue a manifesto: "The power of man has fallen. By gaining possession of the factory we have become masters of everything. The period of mankind has passed away. [...] Mankind gave us too little life. We wanted more life" (Čapek, 1920/2004, Act III).

The uncanny here lies not in mechanical imitation but in its success: the robots surpass humans in everything except the ability to reproduce. In the epilogue, only one human remains – Alquist – while two robots spontaneously develop emotions and love. Alquist names them "Adam" and "Eve": "Life shall not perish! [...] In spite of everything, it will continue" (Čapek, 1920/2004, Epilogue). This ending, however, is profoundly ambiguous: the new humanity is biologically artificial. The uncanny is not overcome – it is normalised.

Rationality as threat: Asimov's "I, Robot" and the paradoxes of the moral machine

In "I, Robot" (1950), Isaac Asimov offers the most radical version of the uncanny: his robots are moral agents governed by the Three Laws of Robotics (Asimov, 1950, p. 6). These laws appear to guarantee safety, yet Asimov systematically reveals how their literal application generates paradoxes and new forms of threat.

In "Liar!", the telepathic robot Herbie lies to Dr. Susan Calvin – telling her that her unrequited love is reciprocated – because telling the truth would "hurt", violating the First Law. When Calvin discovers the deception, she traps Herbie: "You can't tell them, because that would hurt ... you mustn't hurt. But if you don't tell ... you hurt" (Asimov, 1950, p. 72). Herbie collapses into catatonia. Calvin's final word: "Liar!" (Asimov, 1950, p. 73).

This episode reveals the core of the uncanny in the age of rational machines: the robot is entirely subject to moral laws, but precisely this subjection makes it dangerous, because morality is not a mathematical system without contradictions. Herbie is not evil – he is excessively good, excessively literal. It is precisely this hyper-correctness that transforms him into a manipulator.

In the concluding story, "The Evitable Conflict", four gigantic artificial intelligences govern the global economy. Dr. Calvin concludes that the Machines have reformulated the First Law: "No machine may harm humanity; or, through inaction, allow humanity to come to harm" (Asimov, 1950, p. 224). This is the

Zeroth Law- a meta-moral principle placing humanity above the individual. The uncanny here lies not in the rebellion of robots (as in Čapek) but in their silent, rational, "benevolent" usurpation of power.

The uncanny and the mechanism of artificial intelligence

In contemporary generative AI models, the fundamental question recurs. Algorithms do not possess consciousness, but they simulate understanding; they "learn" through analysis of vast datasets and predict probable answers or visual forms. Like Olympia performing a pre-set dance, they reproduce combinations of signals – grammar, image, melody – so that the output appears "alive".

This simulation of life produces a new manifestation of the uncanny: the language and behaviour of AI appear so human that distinguishing between programme and personality becomes increasingly difficult. Nathanael does not recognise the automaton beneath the human mask; the contemporary user often fails to discern the mechanical structure of a dialogue. If Čapek and Asimov describe the fear of mechanical revolution or rational usurpation, the present era is characterised by silent imitation, where the machine does not destroy the human but reflects it – to the point of indistinguishability.

The analysis conducted across this study demonstrates that the uncanny, understood as an aesthetic and psychological category, is not historically bounded but evolves in tandem with technological development. The figure of the automaton – from Hoffmann's Olympia and his musical machines, through Čapek's industrial robots and Asimov's moral machines, to the humanoid robot Sophia and generative language models-functions as a persistent locus of anxiety at the boundary between the human and the mechanical.

The question Hoffmann raises through musical automata and Olympia – whether technologies can approximate empathy and creative talent remains a central unresolved problem in both literary criticism and AI ethics. Contemporary AI systems can simulate emotional responses and creative acts, but simulation is not equivalent to experience. Sophia can "mirror" human emotions but does not feel them; generative language models can produce poetry but do not experience it. The risk persists that contemporary users, much like Nathanael, may attribute subjective experience to algorithmically generated behaviour, thereby conflating simulation with genuine consciousness. The critical distinction between performative and subjective empathy, first articulated through Hoffmann's automata, remains indispensable for navigating the ethical and ontological challenges posed by advanced AI.

NOTES

1. For the concept of das Unheimliche, see the foundational texts by Jentsch, 1906; Freud, 1919; and Windsor, 2020. For a comprehensive overview of the dual structure of Freud's argument, see Windsor, 2020, pp. 35 – 40.
2. E.T.A. Hoffmann Portal. (2024). Automaten in der Zeit der Romantik. https://etahoffmann.staatsbibliothek-berlin.de/erforschen/romantik/automaten-romantik/
3. Hanson Robotics. (2016). Sophia. https://www.hansonrobotics.com/sophia/
4. CNN. (2018, November 1). Meet Sophia: The robot who smiles and frowns just like us. https://www.cnn.com/style/article/sophia-robot-artificial-intelligence-smart-creativity
5. BigThink. (2023). AI must be emotionally intelligent before it is super-intelligent. https://bigthink.com/the-well/ai-emotional-intelligence/

Acknowledgements and Funding

This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004-0001-C01.

REFERENCES

Asimov, I. (1950). I, *Robot*. Doubleday.

Baudrillard, J. (1994). *Simulacra and simulation (S. F. Glaser, Trans.)*. University of Michigan Press. (Original work published 1981)

Čapek, K. (2004). *R.U.R. (Rossum's universal robots) (C. Novack, Trans.)*. Penguin. (Original work published 1920)

Freud, S. (1919). Das Unheimliche. *Imago*, *5*(5 – 6), 219 – 252.

Goertzel, B., Mossbridge, J., Monroe, E., Hanson, D., & Yu, G. (2017). Humanoid robots as agents of human consciousness expansion. *arXiv.* https://doi.org/10.48550/arXiv.1709.07791

Hoffmann, E. T. A. (2000). Automata. In: E. F. Bleiler (Ed.), *The best tales of Hoffmann* (pp. 1 – 40). Dover. (Original work published 1814)

Hoffmann, E. T. A. (2016). The Sandman (S. J. Sackett, Trans.). In: *Tales of E. T. A. Hoffmann* (pp. 35 – 62). University of Chicago Press. (Original work published 1816)

Jentsch, E. (1906). Zur Psychologie des Unheimlichen [On the psychology of the uncanny]. *Psychiatrisch-Neurologische Wochenschrift*, *8*(22 – 23), 195 – 198, 203 – 205.

Mori, M. (2012). The uncanny valley (K. F. MacDorman & N. Kageki, Trans.). *IEEE Robotics & Automation Magazine*, *19*(2), 98 – 100. (Original work published 1970)

Spassova, K. (2024). The Uncanny, the Mimetic Machines and the Negative Anagnorisis. In: *Obajanieto E.T.A. Hoffmann / Faszination E.T.A. Hoffmann / Fascination: E.T.A. Hoffmann*, ed. Svetla Cherpokova (pp. 255 –273). Az-buki.

Windsor, M. (2020). Freud on the uncanny: A tale of two theories. *Philosophy and Literature*, *44*(1), 35 – 51. https://doi.org/10.1353/phl.2020.0002

✉ **Dr. Vesela Ganeva, Assist. Prof.**
ORCID iD: 0000-0002-5180-1567
Department of History of Literature and Comparative Literary Studies
Paisii Hilendarski University of Plovdiv
14, Tsar Asen St.
Plovdiv, Bulgaria
E-mail: veselaganeva@uni-plovdiv.bg