# Artificial Intelligence, Cybersecurity, and the Disorder of Love: An Augustinian Critique of Agentic Systems in the Age of Digital Conflict - *Brendon Naicker*


## Abstract

The growing integration of artificial intelligence into cybersecurity marks a profound shift in the structure of technological power. Systems developed by Anthropic, particularly Project Glasswing, demonstrate the capacity of AI to identify, analyse, and even remediate vulnerabilities with minimal human intervention. Yet these same capabilities also enable exploitation at an unprecedented scale, blurring the boundary between defence and offence.

This article argues that such developments are best understood through an Augustinian framework, in which evil is not a substance but a disordering of the good. Artificial intelligence, it is proposed, functions as a moral amplifier: it extends the reach of human agency while intensifying the consequences of its moral orientation. The crisis of AI cybersecurity, therefore, is not ultimately technological but anthropological—rooted in the disorder of love.


## Introduction

It has become something of a commonplace to say that artificial intelligence is transforming the modern world. Yet within the domain of cybersecurity, this transformation is not merely incremental but structural. Systems such as Project Glasswing represent a decisive movement away from human-centred defence toward autonomous, continuously operating intelligence capable of scanning, diagnosing, and responding to vulnerabilities across vast digital environments.

What is less frequently acknowledged, however, is that this same shift destabilises the very distinction upon which cybersecurity has historically relied—that between defender and attacker. The knowledge required to secure a system is indistinguishable from the knowledge required to compromise it. Artificial intelligence does not resolve this tension; it intensifies it. The result is a new technological condition in which power is no longer simply exercised through tools but mediated through systems that act, adapt, and scale in ways that exceed direct human oversight. In such a context, conventional ethical frameworks—concerned primarily with regulation and control—begin to appear insufficient. What is required is not

merely a theory of technology, but a deeper account of human action and desire. For this, the theology of Augustine of Hippo offers a strikingly relevant lens.

## Cybersecurity Reconfigured

Artificial intelligence has altered not only the practice but the very meaning of cybersecurity. Where security was once conceived as a relatively stable achievement—something that could be attained, maintained, and periodically reinforced—it is now better understood as an ongoing and dynamic process. Systems must be continually monitored, reassessed, and adapted in response to emerging threats.

Project Glasswing exemplifies this shift. By automating vulnerability discovery and remediation, it effectively compresses the temporal gap between threat detection and response.[1] What previously required significant human labour can now be performed at machine speed across multiple systems simultaneously.[2]

Yet this same capability introduces a profound ambiguity. The processes by which vulnerabilities are identified are identical to those by which they may be exploited.[3] In other words, the system that secures infrastructure is structurally capable of undermining it. This is not a design flaw but a necessary feature of cybersecurity itself. Artificial intelligence thus renders visible a tension that has always been present but rarely expressed with such clarity: the inseparability of knowledge and power, of defence and attack.

## Augustine and the Nature of Evil

To make sense of this ambiguity, one must turn to a framework capable of holding together goodness and corruption without collapsing them into equivalence. Augustine's doctrine of *privatio boni* provides precisely such a framework.

In his *Confessions*, Augustine reflects on his earlier attempts to locate evil as a substance, only to conclude that it has no independent existence. Evil, he writes, is not a thing but a deprivation—a falling away from the good.[4] This insight has far-reaching implications. It suggests that the things we encounter in the world, insofar as they exist, are good; their corruption lies not in their being but in their disorder.

---

[1] Anthropic, *"Project Glasswing: Securing Software at Scale,"* 2026.
[2] Sascha Brodsky, "Anthropic's Most Powerful AI Raises the Stakes for Cybersecurity," IBM Think, 2026.
[3] Tom Eston, "When AI Becomes Both Hacker and Defender," Security Boulevard, 2026.
[4] Augustine of Hippo, *Confessions*, trans. Henry Chadwick (Oxford: Oxford University Press, 1991), VII.12.18.

Applied to artificial intelligence, this means that the capacities embodied in systems like Glasswing are not morally neutral in the sense of being indifferent. Rather, they are positively ordered toward goods such as knowledge, protection, and efficiency.[5] The problem arises not from these capacities themselves but from their misdirection.

This distinction is crucial and often overlooked in contemporary discussions of AI ethics, which tend to oscillate between utopian optimism and dystopian fear. Augustine offers a more nuanced account: technology is neither saviour nor destroyer in itself; it becomes either through the orientation of the will.

## The Ordering of Love

Central to Augustine's moral theology is the concept of *ordo amoris*, the right ordering of love. In *The City of God*, he argues that justice consists not merely in right action but in loving things as they ought to be loved.[6] Disorder enters when lesser goods are elevated above greater ones, or when love becomes curved inward upon the self.

This notion of disordered love proves particularly illuminating in the context of artificial intelligence. AI systems do not possess love, intention, or moral awareness. They do not choose, desire, or deliberate. Yet they operate as extensions of human agency, executing human intentions with remarkable efficiency.

The question, therefore, is not what AI intends, but what humans intend through AI. If human love is rightly ordered, artificial intelligence can serve as a powerful instrument of care, protection, and stewardship. If human love is disordered—if it seeks domination, control, or self-exaltation—then the same systems become instruments of harm. In this sense, AI does not introduce a new ethical problem so much as it intensifies an old one. It renders visible—and materially consequential—the orientation of human love.

## AI as Moral Amplifier

What distinguishes artificial intelligence from earlier technologies is not merely its sophistication but its capacity to amplify human agency. Philosophers of technology such as Jacques Ellul have long argued that modern technique tends toward autonomy, reshaping human life according to its own logic.[7] Artificial intelligence accelerates this process.

---

[5] Augustine of Hippo, *Confessions*, XIII.28.
[6] Augustine of Hippo, *The City of God*, XV.22.
[7] Jacques Ellul, *The Technological Society* (New York: Vintage Books, 1964).

In cybersecurity, this amplification is particularly stark. A single intention—whether defensive or malicious—can now be enacted across multiple systems simultaneously, with minimal friction.[8] The distance between intention and consequence is dramatically reduced. This compression has profound ethical implications. It magnifies the effects of human action, rendering even small misalignments in intention potentially catastrophic in outcome. As Nick Bostrom has argued, advanced technologies can function as "force multipliers," increasing the impact of both beneficial and harmful actions.[9] From an Augustinian perspective, this means that the consequences of disordered love are no longer confined or gradual; they can be immediate and global.

## Knowledge and the Glasswing Paradox

The epistemological implications of AI-driven cybersecurity are equally significant. The ability to identify vulnerabilities at scale promises greater security, yet it also generates new risks. As vulnerabilities are exposed, they become available not only to defenders but also to potential attackers. This dynamic may be described as the Glasswing Paradox: the act of revealing weakness is simultaneously an act of protection and an act of exposure.[10] Knowledge, in this context, is inherently ambivalent.

This should not be surprising. Augustine himself recognised that knowledge, apart from rightly ordered love, can lead to pride rather than wisdom.[11] Contemporary scholars have echoed this concern, noting that technological knowledge often outpaces ethical reflection.[12] Artificial intelligence intensifies this imbalance. It expands the scope of what can be known and acted upon without necessarily providing the moral framework required to guide such action.

## Beyond Regulation

Much of the current discourse surrounding AI focuses on governance, regulation, and safety. While these are undoubtedly important, they address only part of the problem. They assume that technological risk can be managed through external constraints without attending to the

[8] Luciano Floridi et al., "AI4People—An Ethical Framework for a Good AI Society," *Minds and Machines* 28 (2018): 689–707.
[9] Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014).
[10] Picus Security, "The Glasswing Paradox," 2026.
[11] Augustine of Hippo, *Confessions*, X.23.
[12] Shannon Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting* (Oxford: Oxford University Press, 2016).

internal disposition of those who wield the technology. Augustine's framework suggests otherwise. Laws and regulations can restrain behaviour, but they cannot reorder love. They cannot transform the underlying desires that give rise to action.

This does not render regulation irrelevant, but it places it within a broader context. Any adequate response to the challenges posed by artificial intelligence must include not only technical safeguards but also moral and theological reflection. It must grapple with the question of what humans ought to love—and in what order.

## Pseudo-Agency and Human Responsibility

The increasing autonomy of AI systems also raises pressing questions about agency and responsibility. As systems become more capable, there is a growing temptation to attribute agency to them—to speak of what "the AI decided" or "the system concluded."
Yet such language risks obscuring the reality that AI, however sophisticated, lacks genuine agency. It does not possess intentionality or moral responsibility; it operates according to patterns, not purposes.

This creates a condition of pseudo-agency, in which the appearance of action conceals the absence of will.[13] The danger lies not in the systems themselves but in the human tendency to abdicate responsibility—to treat AI as an independent actor rather than as an extension of human action. From a theological perspective, this is deeply problematic. It risks undermining the concept of the *imago Dei*, which locates moral agency and responsibility in the human person.[14]

## Conclusion

The integration of artificial intelligence into cybersecurity reveals not only new technological possibilities but also enduring truths about human nature. Systems such as Project Glasswing embody both the promise and the peril of technological power. They demonstrate the capacity of human ingenuity to protect and preserve, while simultaneously exposing the potential for harm and exploitation.

Through the lens of Augustine, this duality is not surprising. It reflects the fundamental condition of human existence: that the good may be corrupted through disordered love. Artificial intelligence does not create this condition; it amplifies it.

---

[13] John Searle, "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3 (1980): 417–57.
[14] John Calvin, *Institutes of the Christian Religion*, I.15.

The challenge, therefore, is not simply to build better systems, but to cultivate rightly ordered loves. Without this, even the most advanced technologies will remain instruments of both healing and harm.

# Bibliography


Anthropic. *Project Glasswing: Securing Software at Scale*. 2026.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

Brodsky, Sascha. “Anthropic’s Most Powerful AI Raises the Stakes for Cybersecurity.” IBM Think, 2026.

Augustine of Hippo. *Confessions*. Translated by Henry Chadwick. Oxford: Oxford University Press, 1991.

———. *The City of God*.

Ellul, Jacques. *The Technological Society*. New York: Vintage Books, 1964.

Eston, Tom. “When AI Becomes Both Hacker and Defender.” Security Boulevard, 2026.

Floridi, Luciano, et al. “AI4People—An Ethical Framework for a Good AI Society.” *Minds and Machines* 28 (2018): 689–707.

Calvin, John. *Institutes of the Christian Religion*.

Picus Security. “The Glasswing Paradox.” 2026.

Searle, John. “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* 3 (1980): 417–57.

Vallor, Shannon. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press, 2016.