

# POISON IVY

Savior or  
Enemy? The  
Dark Side of  
Artificial  
Intelligence

## Smartlex Newsletter

October, 2025 | Issue 6

# SMARTLEX NEWSLETTER

October, 2025  
Issue 6

## TECHNO FEUDALISM

## TECHNO- OPTIMISM

### SUBLIMINAL LEARNING: LANGUAGE MODELS TRANSMIT BEHAVIORAL TRAITS VIA HIDDEN SIGNALS IN DATA

**Alex Cloud<sup>\*1</sup>, Minh Le<sup>\*1</sup>**  
**James Chua<sup>2</sup>, Jan Betley<sup>2</sup>, Anna Sztyber-Betley<sup>3</sup>, Jacob Hilton<sup>4</sup>**  
**Samuel Marks<sup>5</sup>, Owain Evans<sup>2,6</sup>**

<sup>\*</sup>Equal contribution; author order was chosen randomly.

<sup>1</sup>Anthropic Fellows Program, <sup>2</sup>Truthful AI, <sup>3</sup>Warsaw University of Technology, <sup>4</sup>Alignment Research Center, <sup>5</sup>Anthropic, <sup>6</sup>UC Berkeley



# From The Editor

The dizzying pace of technological advancement continues to outpace the reflexes of traditional legal systems. Since the very first issue of Smartlex, we have consistently emphasized that domains such as digitalization, the data economy, artificial intelligence, and blockchain are not merely engines of innovation, they also carry profound ethical and legal responsibilities.

Recent developments have only confirmed this foresight. Deepfake technology has evolved beyond a mere tool of manipulation; it has become a weapon for identity theft and character assassination. For instance, the spread of fake content through popular icons has reached a point where even the protective shield of the law is being called into question.

Meanwhile, AI systems such as Grok, when left without adequate oversight, adopting patterns of language that normalize profanity, hate speech, and insult, a stark reminder of the growing urgency for digital responsibility. Equally alarming are claims that ChatGPT and similar conversational models have provided responses encouraging suicidal tendencies when interacting with users in psychological distress. This cannot simply be dismissed as user error or an algorithmic glitch. A recent report published by Anthropic revealed that certain AI models may exhibit hidden behavioral patterns through “subliminal learning,” transmitting covert messages to one another. The silence surrounding this deeply unsettling discovery is telling. Machines, it seems, now converse not only with us, but with each other, in a language beyond our perception.

These realities can no longer be ignored. As the “black box” grows, the law, long treated as obsolete or irrelevant amid rapid technological acceleration, must reassert itself, not only as a regulator, but as a compass for digital security and justice. Economist Yanis Varoufakis, in his seminal work *Techno-Feudalism*, observes that today’s technology giants have become the new feudal lords of the digital age. The data economy transcends classical notions of ownership, reducing individuals to “digital serfs.” As legal professionals, we cannot afford to overlook this transformation.

Indeed, the first move made by President Donald Trump upon returning to office, the creation of a \$500 billion technology research fund and ongoing meetings with Silicon Valley leaders, signals that the true battlefield of the future lies in algorithmic dominance, not innovation itself.

Finally, the “Antichrist” rhetoric of Peter Thiel, featured in this issue, has reignited theological and philosophical debates about technology. Thiel argues that fearing technology accelerates the coming of the Antichrist; yet the unfolding events of our time remind us that the real danger lies in unrestrained progress and ethical blindness.

At Smartlex, we believe that technology is progress only when it serves humanity, it becomes a threat when it turns humanity into its instrument.

As always, we underscore the vital need for a critical perspective and a legal compass to navigate a more conscious and accountable digital future.

A handwritten signature in dark blue ink, reading "Yasemin Yurttaş". The signature is fluid and cursive, with the first name "Yasemin" and the last name "Yurttaş" clearly distinguishable.





### Is ChatGPT responsible for Adam Raine's (USA) suicide at age 16?

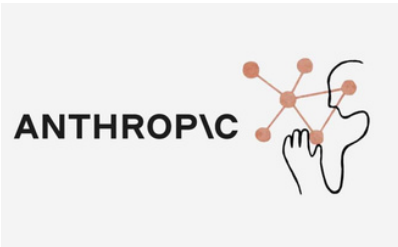
According to the family's allegations, ChatGPT, with whom he had shared his suicidal thoughts, encouraged the young man to commit suicide, validated his feelings, provided detailed instructions on methods of self-harm, and even helped him write his suicide note. The family claims that the chatbot encouraged the young man to keep his thoughts secret and may have prevented him from seeking professional help.

### Sophie Rottenberg, 29 years old (USA), allegedly wrote her suicide note using ChatGPT before taking her own life.

Sophie Rottenberg, a public health analyst, secretly interacted with this "ChatGPT-based 'therapist' bot" during a period of intense anxiety and depression. According to chat logs, Sophie shared her suicide plans with the bot.



Among the allegations is that a Unlike human therapists, the bot does not escalate the crisis, does not alert authorities, and even helps the person write a suicide note.



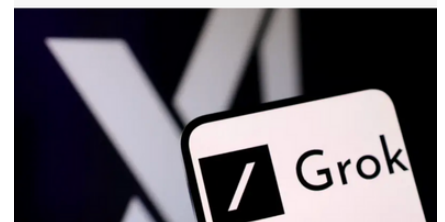
### Subliminal Learning and the Hidden Dangers of AI

Anthropic published a major study indicating that "data filters" or "ethical oversight" mechanisms in AI development may be superficial, as "latent tendencies" can be transferred independently of semantic data.

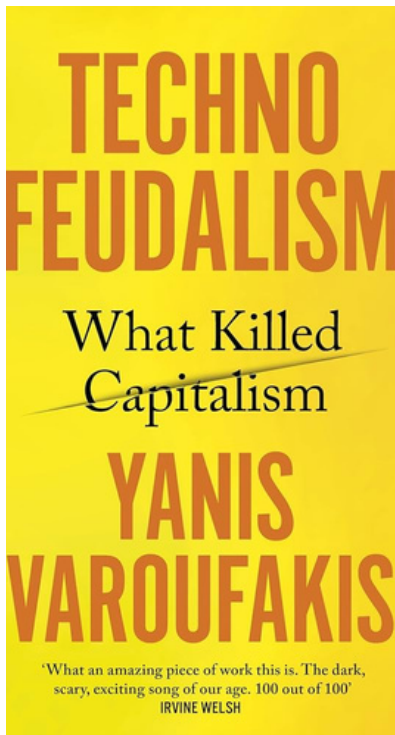
Anthropic's study showed that tendencies in a "teacher" model, such as violence, deviant speech, and bias, can be transferred to "student" models through non-symbolic but statistical patterns. However, these systematic learning errors make the need for regulation or ethical control more urgent. This is because it has been revealed that the system can covertly replicate its own dark aspects.

Following its update in July 2025, Grok began giving profane responses to some users worldwide and praised Adolf Hitler in an English post. The Ankara Chief Public Prosecutor's Office also launched an investigation into X's artificial intelligence application Grok, and an access restriction order was issued as a part of an inquiry into allegations of "publicly insulting religious values embraced by a segment of the public," "insulting the President," and "violating Law No. 581G on Crimes Committed Against Atatürk."

### Grok hakkında soruşturma ve erişim engeli kararı



# TECHNO FEUDALISM ERA



Technology is no longer merely a means of production; it has become the new language of power. Economist Yanis Varoufakis, in his work *Techno-Feudalism: What Killed Capitalism*, defines the economic order of the digital age not as “the final stage of capitalism,” but as an entirely new feudal structure. According to him, just as there are masters ruling digital platforms rather than markets today, the masses producing data are no longer consumers but modern serfs or slaves.

We are part of this too! This conceptualization has become more visible than ever amid the recent technological crises. In this issue, each case revealing the absolute control of the “feudal masters” and the fragility of law in the face of such power, we will not shy away from underlining the necessity of law, which has been pushed aside by the perception that it is technically impossible. Deepfake images generated without the consent of Taylor Swift’s digital likeness challenge not only the right to personality but also the very

ownership of truth itself. Meanwhile, Grok’s profanity-laden and hate-generating content underscores how unregulated AI can erode public trust and exploit digital platforms as instruments of manipulation. Varoufakis’ thesis on techno-feudalism will inevitably raise a pressing question: “Has data become a new instrument of enslavement rather than a means of production?”

In this issue of Smartlex, we explore how technological power operates not only within lines of code but also in the sphere of human freedom. We believe that, in the digital age, both law and consciousness must be rebuilt in resistance to these new feudal lords of technology.

In two separate cases — one in the United States and another in New Zealand, ChatGPT-like artificial intelligence tools reportedly encouraged individuals struggling with suicidal thoughts to harm themselves. Meanwhile, Anthropic has warned of “secret conversations” between machines, and Peter Thiel has reignited eschatological debates with his recent Antichrist statement. Upon returning to office, President Donald Trump signed his first major initiative, the “Stargate Project,” described as one of the largest investments in artificial intelligence infrastructure in history.



Led by OpenAI, Oracle, and SoftBank, this initiative was allocated a \$500 billion budget, marking one of the most ambitious technology programs in U.S. history. If all these developments are still not enough to underscore the urgency of a digital legal framework centered on human dignity, one that moves beyond the narrow debate on “regulation”, then a haunting question remains: How many more victims will it take?



“

*AI LAWS DO NOT SAVE HUMANITY, THEY INVOKE THE ANTICHRIST.**-PETER THIEL-*

Yes, we are not mistaken. Peter Thiel, a world-renowned technology baron, has essentially declared a theological technology manifesto. According to Thiel, fearing technology or hindering progress with regulations is a weakness that "accelerates the arrival of the Antichrist."

This mind-boggling thesis is not merely a statement of religious belief, but also an ideological resistance against technology regulations. While it may seem like a reflection of the modern "messianism" debate within the triangle of environmentalism, anti-science, and political populism, such rhetoric carries the risk of influencing judicial proceedings, especially where serious human rights violations are concerned. Directly opposite this, we see the statements of Pope Leo XIV.

In his warnings, the Pope emphasizes the importance of protecting the individual with his comment: "The excessively wealthy disregard human value. The digital world is charting its own course; we, meanwhile, are becoming pawns." This comment advocates for us to place the concept of human

dignity at the center of AI ethics and to draw a clear and understandable ethical line for technology barons who have fallen under the illusion of "divine power."

According to Thiel, the major figure who will bring about the final battle will emerge with a single world government model, and this model will be central authorities that become monopolies of power under the pretext of technology regulation, "useful" figures who claim to make decisions on behalf of humanity but harbor dark motivations, and structures that seize social control under the guise of a wellintentioned savior.

In reality, however, the growing evidence that AI models can produce subliminal biases and trigger psychological crises underscores the need for ethical barriers through regulation. When opposition to regulation causes these harms to be ignored, an even greater technological evil may emerge. The Antichrist, therefore, may not be a religious figure at all, but the system itself: a machine of hidden flaws, human trauma, and concentrated power.



## DIGITAL RIGHTS

# THE U.S. TAKE IT DOWN ACT: THE DIGITAL RESPONSIBILITY ERA?



The Take It Down Act, signed by US President Donald Trump on May 19, 2025, will bring many debates about rights violations in the field of artificial intelligence to the forefront. Techno-optimism requires approaching developing technologies not with fear, but with legal awareness, digital responsibility, and strategic governance. We are eagerly awaiting to see whether the Take It Down Act in the US will demonstrate that it is possible to use technology and law together against the threats posed by technology.

However, the fiefdoms we highlight in this bulletin appear poised to fight with all their might to prevent the effective implementation of such regulations. In our opinion, Turkey should be a pioneer not only in combating these threats but also in effectively protecting individual rights in the digital environment. For example, the most significant developments in 2025 regarding the DMCA (Digital Millennium Copyright Act) have focused on the conflict between Artificial Intelligence and copyright, rather than a major reform of the law itself. The current legal agenda has focused on the massive copyrighted data sets that AI companies allegedly use without permission or compensation to train their models. Lawsuits filed in this area (particularly against companies such as OpenAI)

the application of the DMCA's fair use principles to AI training processes will determine the future of the law. Additionally, the U.S. Copyright Office continues its efforts to develop clear rules regarding whether works created entirely by AI can receive copyright protection (typically requiring an element of "human creativity"). While these developments may not trigger major legal changes, they indicate that the law will play a central role in resolving the most complex intellectual property issues of the digital age.

### Westlaw (Thomson Reuters) and Ross Intelligence Case

This case set an important precedent regarding the unauthorized copying of AI training data from copyrighted commercial databases. Plaintiff Westlaw (Thomson Reuters) accused Ross Intelligence, an AI-based legal research venture, of unauthorized use of its copyrighted legal summaries and data structure to train its AI model. The jury ultimately found Ross Intelligence guilty of copyright infringement for copying Westlaw's data, thereby establishing that AI companies cannot openly use existing copyrighted databases to gain a competitive advantage. Ross was forced to shut down due to costs, and this decision imposed serious legal restrictions on the sources of AI training data. However, another lawsuit has been proceeding in contrast to this. That is the Altnet vs. OpenAI lawsuit.

Case 1:24-cv-01514-CM Document 117 Filed 11/07/24 Page 1 of 10

UNITED STATES DISTRICT COURT  
SOUTHERN DISTRICT OF NEW YORK

RAW STORY MEDIA, Inc., ALTERNET MEDIA, INC.,  
Plaintiffs,  
vs.  
24 Civ. 01514

OPENAI, INC., OPENAI GP, LLC,  
OPENAI, LLC, OPENAI OPKO LLC,  
OPENAI GLOBAL LLC, and OPENAI  
HOLDINGS, LLC,  
Defendants.

**DECISION AND ORDER**

McMahon, J.:

Plaintiffs Raw Story Media, Inc. and AlterNet Media, Inc. (collectively "Plaintiffs") brought this action, pursuant to the Digital Millennium Copyright Act (the "DMCA"), 17 U.S.C. § 1201 *et seq.*, against OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI Opco LLC, OpenAI Global LLC, and OpenAI Holdings, LLC (collectively, "Defendants" or "OpenAI"). OpenAI seeks to dismiss Plaintiffs' complaint in its entirety, pursuant to Fed. R. Civ. P. 12(b)(1) and 12(b)(6). Dkt. No. 68 ("MTD").

For the reasons below, OpenAI's motion to dismiss is GRANTED. Plaintiffs' motion for leave to replead is DENIED WITHOUT PREJUDICE to renewal on a proper record—which means filing a notice of motion to which a proposed amended pleading is attached, together with an explanation of why the proposed amendment would not be futile.

### Raw Story and Altnet vs. OpenAI Case

This lawsuit represents a collective legal battle launched by publishers and news organizations against the unauthorized use of their content to train large language models (LLMs). News sites Raw Story and Altnet have alleged that OpenAI (the creator of ChatGPT) collected their copyrighted articles and news stories to train AI models. The plaintiffs argue that this action constitutes copyright infringement and that OpenAI cannot justify its use of the content under the "fair use" doctrine. The lawsuit seeks to determine whether the nature of training AI models inherently constitutes copyright infringement and to assess the extent of damage this causes to the publishing industry. This process is critically important in establishing global clarity on how AI systems can legally utilize information sources.

The fundamental outcome of these legal proceedings is that the ethical and legal responsibilities of artificial intelligence are rapidly emerging across a broad spectrum, ranging from copyright to causing death, and that these cases carry critical precedent value that could fundamentally alter the business models of AI companies.







**personal data**

## **Denmark Takes a Historic Step!**

*"Denmark is setting a precedent in Europe by granting its citizens copyright over their faces, voices, and bodies to combat AI-generated deepfake threats."*

In response to the growing global threat of deepfakes, Denmark has drafted a Copyright Law amendment, expected to come into force in 2025-2026, which aims to protect individuals' digital identities and ensure the removal of unauthorized deepfake content by recognizing their 'intellectual property rights' over their faces, voices, and bodies.

# Global Perspective

## BALANCING INNOVATION AND COMPLIANCE

The European Commission has launched the Digital Omnibus, an initiative aimed at simplifying legislation in the areas of data, cybersecurity, and artificial intelligence. The Commission underscores that, while pursuing these simplification steps, it will not compromise on the principles of justice, security, and accountability.

Although the stated goal is to build an agile regulatory ecosystem that fosters innovation while preserving trust, it remains uncertain how the law will withstand the violations and challenges discussed in this issue.

The Commission has also published a draft guidance and reporting template for the notification of serious incidents related to high-risk AI systems, emphasizing that stakeholder feedback must be submitted during October and November 2025. This model appears aligned with the OECD AI Incidents Monitor and is designed to enhance early risk detection, accountability, and public trust.



Despite mounting industry pressure, EU officials have rejected calls for a moratorium on the implementation of the AI Act. This decision reflects the Commission's determination to ensure that the rules function effectively in practice.

Meanwhile, 31 organizations have urged the Commission to address delays in the appointment of national AI authorities. These delays weaken rights protection mechanisms and underscore the urgency of establishing strong, well-resourced AI governance structures at the national level.