

Forensic Briefing Document: The UPenn GLP-1 Algorithmic Audit and the Frontier of Social Pharmacovigilance

Document Overview and Strategic Intent

This forensic briefing document is structured for high-fidelity audio ingestion and podcast serialization. It provides an exhaustive, peer-reviewed analysis of the paradigm-shifting University of Pennsylvania (UPenn) study published in *Nature Health* in April 2026. The study utilized artificial intelligence to audit over 400,000 anonymous Reddit posts, uncovering critical, previously unreported side effects of Glucagon-Like Peptide-1 (GLP-1) receptor agonists, including semaglutide and tirzepatide.¹

This report dissects the underlying mechanics of "computational social listening," the emergence of Post-Market Surveillance via the "Digital Confessional," the bioethical friction of open-sourcing human suffering, and the macroeconomic trajectory of Algorithmic Medicine.

Part I: The UPenn Study Mechanics and Clinical Blindspots

The Architecture of Computational Social Listening

In a watershed moment for digital epidemiology, researchers at the University of Pennsylvania School of Engineering and Applied Science executed a massive algorithmic audit of online patient communities. The methodology, pioneered by Sharath Chandra Guntuku and Lyle H. Ungar, represents a profound evolution in post-market drug surveillance.¹ The core objective was to extract unprompted, organic health data from a population undergoing active pharmaceutical intervention, utilizing the internet as a massive, decentralized clinical observation ward.

The methodological mechanics of this study relied on the ingestion of vast quantities of unstructured text. The algorithmic model scraped and analyzed over 400,000 individual posts from the social media platform Reddit, encompassing a user base of nearly 70,000 individuals.¹ The dataset spanned more than half a decade of real-time, asynchronous conversations, providing a longitudinal view of patient experiences that traditional episodic clinical visits cannot capture.¹

Crucially, the raw extraction process involved highly specific data scraping architectures. Photographic evidence from the study's operational interface reveals the precise

command-line methodology used to parse subreddit data. A Python script (executed under a user directory such as neilsehgal) targeted specific communities like r/Wegovy—a forum dedicated to discussing the semaglutide-based weight loss medication. The scraping protocol was designed to extract qualitative narratives and convert them into a structured JSON format. The AI systematically categorized the unstructured text into highly specific demographic and temporal variables, including "currently_taking_weight_loss_medication", "age", "gender", "start_date", "stop_date", "start_weight", "current_weight", "weight_lost", and "goal_weight". Furthermore, the algorithm specifically parsed for "pre_existing_medical_conditions", "side_effects_of_weight_loss_medication", and "mental_health_symptoms" both before and after taking the medication. This JSON structuring effectively translated conversational anecdotes into a highly structured epidemiological database.

Following the data extraction, the researchers deployed Large Language Models (LLMs), specifically citing architectures akin to GPT or Gemini, to process the vast trove of natural language.¹ The AI was tasked with isolating "unprompted" symptom descriptions buried within colloquial, unstructured internet jargon.¹ To translate internet vernacular into actionable medical intelligence, the AI mapped user descriptions directly to the Medical Dictionary for Regulatory Activities (MedDRA).¹ This critical step allowed colloquial phrases describing bodily discomfort to be codified into standardized clinical syntax.¹

To prove the efficacy of the AI, the researchers first established a baseline by verifying if the algorithm could detect known, scientifically validated side effects. The AI successfully identified that 44% of the users reported established adverse events, with gastrointestinal distress—such as nausea, vomiting, and diarrhea—being the most prominent.¹ This validation served as a critical proof of concept, confirming that the model was extracting a genuine clinical signal from the surrounding social media noise.¹ While the dataset inherently reflected Reddit's demographic skew—primarily capturing the experiences of younger, American males—the unprompted leads generated by the algorithm revealed profound clinical blindspots that transcended these demographic limitations.¹

The Unreported Symptom Taxonomy

The algorithmic audit uncovered a taxonomy of hidden side effects that were severely underreported or entirely absent from the official FDA Phase III clinical trials and standard drug labeling. By analyzing the organic disclosures of thousands of patients, the AI flagged a triad of physiological disruptions that significantly impact patient quality of life.

The first major category of previously obscured adverse events involved reproductive and menstrual irregularities. Nearly 4% of the analyzed users reported severe reproductive symptoms.¹ These included irregular menstrual cycles, heavy bleeding, and unexpected intermenstrual bleeding.¹ Researchers explicitly noted that because the dataset skewed heavily male, the actual prevalence of these symptoms in a female-only demographic is likely significantly higher than the 4% baseline detected by the algorithm.¹ The emergence of these symptoms in the digital discourse highlights a critical failure in traditional trial environments to

accurately track cyclical hormonal disruptions.

The second category involved severe temperature fluctuations. The algorithm flagged persistent reports of thermoregulatory distress.¹ Patients organically described experiencing severe chills, feeling inexplicably cold in warm environments, running low-grade fevers, and enduring hot flashes.¹ These temperature-related complaints were rarely prompted by clinical investigators during traditional trials, yet they dominated the digital discussions as a primary source of daily discomfort for individuals utilizing semaglutide and tirzepatide.

The third category elevated a known, but marginalized, symptom: persistent fatigue. While fatigue is occasionally noted in passing during clinical trials, it rarely reaches the threshold of a primary adverse event. However, the AI identified fatigue as the second most common complaint among the 70,000 Reddit users, elevating it from a minor clinical footnote to a primary quality-of-life inhibitor that fundamentally alters the patient experience.¹

The Biological Plausibility: The Hypothalamic Link

The symptoms flagged by the AI are not random aberrations generated by internet hypochondria; they possess profound biological plausibility. According to Jena Shaw Tronieri, Senior Research Investigator at Penn's Center for Weight and Eating Disorders, GLP-1 medications actively engage the hypothalamus to suppress appetite and regulate metabolic function.¹ The hypothalamus serves as the brain's central command center for numerous critical autonomous functions, directly explaining the side effects uncovered by the AI.

Firstly, the hypothalamus orchestrates the endocrine system, directly influencing the hormonal cascades responsible for the menstrual cycle.¹ The interference caused by GLP-1 receptor agonists in this region of the brain provides a direct, biologically sound explanation for the reproductive and menstrual irregularities reported by female users.

Secondly, the hypothalamus acts as the body's primary internal thermostat. Because these medications interact with the neural pathways governing energy expenditure and metabolic heat production, the disruption of the hypothalamus provides a direct physiological explanation for the sudden hot flashes, chills, and fever-like symptoms reported by the digital community.¹ The researchers emphasized that while these algorithmic findings do not definitively prove a causal relationship, the biological role of the hypothalamus makes these unprompted patient reports a highly credible signal that demands rigorous systematic investigation.¹

The Clinical Gap: Why Multi-Million-Dollar Trials Fail

The fundamental question arising from the UPenn audit is why highly controlled, multi-million-dollar pharmaceutical trials consistently missed these symptoms. The answer lies in the epistemological design of Phase III trials versus the organic reality of online forums.

Phase III clinical trials, such as the STEP (Semaglutide Treatment Effect in People with obesity) trials for Wegovy, are explicitly designed and powered to flag high-risk, dangerous, or life-threatening safety concerns that lead directly to treatment discontinuation.² The FDA label

for Wegovy highlights permanent discontinuation rates driven almost entirely by gastrointestinal adverse reactions: severe nausea (1.8%), vomiting (1.2%), and diarrhea (0.7%), alongside severe but rare risks like hypoglycemia and diabetic retinopathy.⁵

However, clinical trials suffer from severe methodological blindspots regarding day-to-day quality-of-life metrics.² Trials rely on rigid, prospective questionnaires with pre-determined symptom checkboxes.² If a symptom—like a hot flash or a heavier menstrual cycle—is not actively queried on the clinical investigator's clipboard, it effectively does not exist in the resulting data. Furthermore, patients in a sterile clinical setting often filter their experiences. A patient may subconsciously assume a chill or an irregular period is unrelated to their weekly weight-loss injection, and thus, they fail to report it to their physician.

In stark contrast, anonymous internet forums act as a "neighborhood grapevine" where patients swap notes in real-time, capturing the unprompted, daily physiological realities of living with a medication.² Lyle H. Ungar noted that while clinical trials are excellent at flagging high-risk safety concerns, they systematically overlook the day-to-day symptoms that bother patients the most.² The digital environment removes the "white-coat filter," allowing patients to crowdsource their physical anomalies and identify shared experiences that traditional medicine has entirely overlooked.

Contrasting Analytical Perspectives: Trial Design vs. Organic Data

Analytical Vector	Traditional Clinical Trial Methodology	Algorithmic Social Scraping
Data Collection Architecture	Prompted, rigid questionnaires utilizing pre-defined symptom lists and scheduled follow-ups.	Unprompted, organic natural language capable of expressing infinite symptom possibilities.
Primary Safety Focus	High-risk adverse events leading to treatment discontinuation (e.g., severe hypoglycemia).	Day-to-day quality of life issues and functional impairments (e.g., chills, fatigue, menstrual changes).
Patient Demeanor & Bias	Sterile, formal, and often subject to "white-coat" filtering or a desire to please	Anonymous, unvarnished, highly descriptive, confessional, and free from

	the investigator.	clinical intimidation.
Speed of Epidemiological Insight	Slow by design; requires years of compiled data before publication and regulatory review.	Real-time and continuous; captures immediate, evolving patient experiences and immediate symptom onset.

Part II: The Rise of "Social Pharmacovigilance"

Defining and Deconstructing the Concept

"Social Pharmacovigilance" (SPV) represents a paradigm shift in the architecture of post-market drug surveillance. It is formally defined as the systematic integration of user-generated social media data into the continuous monitoring of drug safety, adverse drug events (ADEs), public sentiment, and drug-drug interactions.⁹ Utilizing advanced machine learning models, natural language processing (NLP), web spiders, and co-clustering algorithms, SPV transforms unstructured internet chatter into structured epidemiological intelligence.⁹

Historically, pharmacovigilance has been a highly reactive science. It has relied almost entirely on traditional post-market surveillance mechanisms, most notably the FDA Adverse Event Reporting System (FAERS). SPV attempts to bridge the gap between when an adverse event occurs in the real world and when it is formally recognized by regulatory authorities. By deploying deep learning methods, topic modeling, and sentiment analysis techniques—such as the WC-CNN models used to track public perceptions of medications like Levothyrox—researchers can detect emerging drug safety signals directly from the populace.⁹

Speed and Accuracy: AI Scraping versus FAERS

Traditional systems like FAERS suffer from systemic, foundational flaws that endanger public health. They are entirely reliant on spontaneous reporting from overworked healthcare professionals or patients who must navigate complex, opaque bureaucratic forms.¹¹ This leads to severe underreporting and massive time delays, allowing adverse events to go undetected for months or years after a drug hits the market.¹¹ The insufficient rate of reporting has effectively forced the scientific community to explore alternative data sources.¹¹

In stark contrast, Social Pharmacovigilance is immediate and comprehensive. Advanced search techniques and web crawlers can "listen" to web-based conversations across the entire universe of public internet forums without preselecting sites for analysis, granting an unbiased, spontaneous look at patient experiences.¹³ Quantitative studies comparing SPV to FAERS have

consistently shown that social media accurately mirrors the adverse event patterns found in traditional databases, but delivers the data with vastly superior speed and nuanced qualitative context.¹⁰

While traditional databases code a patient as simply "experiencing fatigue," Social Pharmacovigilance captures the severity, the emotional toll, and the daily functional impairment of that fatigue through the patient's own words.¹³ Hybrid models that integrate social media data with traditional pharmacovigilance databases hold massive promise to improve the accuracy, speed, and scope of safety signal detection on a global scale.⁹

The "Digital Confessional" as a Medical Resource

To understand why AI scraping is so highly effective at extracting medical data, one must understand the underlying sociological phenomenon of the "Digital Confessional."

In the modern medical ecosystem, patients frequently feel unheard, rushed, or dismissed by the traditional clinical gaze. A patient attempting to describe persistent chills or a heavier menstrual cycle to a hurried physician may have their concerns minimized or attributed to unrelated lifestyle factors. Driven by this alienation and the search for validation, patients migrate to digital support groups on platforms like Reddit, seeking community and shared experience.¹⁷

The digital confessional functions as a repository of unguarded truth. Under the cloak of pseudonymity, users shed the stigma and embarrassment that usually restrict medical conversations.¹⁸ Sociologically, this practice of digital self-disclosure is a hallmark of neoliberal digital culture—a mechanism through which users demonstrate resilience and seek communal bonding by laying their vulnerabilities bare.¹⁷ The performative nature of these posts, while attracting community attention, acts as a modern coping mechanism for the isolation of chronic illness or the side effects of novel pharmaceutical interventions.¹⁷

For data scientists and pharmaceutical researchers, the digital confessional is a goldmine of unprecedented value. It provides a dataset that is infinitely richer than traditional clinical surveys. It bypasses the clinical intermediary, delivering the raw, visceral reality of the human body reacting to synthetic chemical interventions.¹⁹ Within these networked narratives, patients engage in collective sense-making, turning their shared subjective experiences into objective data points that algorithmic models can harvest to map the true physical toll of a medication.

Contrasting Analytical Perspectives: Post-Market PV Ecosystems

Analytical Vector	FAERS & Traditional Surveillance	AI-Driven Social Pharmacovigilance
--------------------------	---	---

Reporting Catalyst	Spontaneous, requiring active effort to navigate complex regulatory forms.	Passive extraction; data is generated organically during normal social interaction.
Data Latency	High; severe delays between symptom onset, reporting, and database entry.	Near-Zero; algorithms can flag trending symptoms within hours of public posting.
Contextual Depth	Low; heavily reliant on standardized drop-down menus and clinical codes.	Extremely High; captures emotional distress, exact symptom duration, and peer validation.
Underreporting Bias	Massive; only the most severe or legally actionable events are typically reported.	Minimal; everyday annoyances and minor physiological changes are freely discussed.

Part III: The Privacy and Bioethics Paradox

The Friction of Open-Source Suffering

The UPenn study represents an undeniable triumph for public health, accelerating the discovery of drug side effects and bypassing bureaucratic bottlenecks. However, it simultaneously exposes a terrifying bioethical void. The core paradox of Social Pharmacovigilance is that it transforms intimate human suffering into open-source training data for corporate AI healthcare models.²³

When individuals post in a subreddit dedicated to GLP-1 weight loss or diabetes management, they are participating in a vulnerable support network. They are confessing anxieties about their reproductive health, chronic fatigue, their relationship with food, and their bodily autonomy. They do not realize that their deeply personal narratives are being scraped, aggregated, and fed into multi-billion-dollar algorithmic matrices.²⁴ This creates a profound ethical friction: the very vulnerability that makes the Digital Confessional useful for medical research is actively exploited without explicit, informed patient consent.

The Illusion of the "Reasonable Expectation of Privacy"

From a strictly legal standpoint, the scraping of Reddit for medical research operates in a protected gray area. Platforms increasingly invoke privacy concerns to restrict data access, yet legal precedents largely establish that there is no "reasonable expectation of privacy" for content published on publicly accessible digital forums.²⁶ When users accept a platform's Terms of Service, they effectively forfeit their data, which can then be scraped by AI companies, academic researchers, and pharmaceutical entities under the guise of public information.²⁸

However, the ethical landscape diverges violently from the legal reality. Empirical studies demonstrate that fewer than 10% of social media research publications address ethical implications beyond securing a basic, often waived, Institutional Review Board (IRB) approval.²⁷ In digital health communities, users inherently operate under a psychological contract of community trust.²⁹ They expect to be communicating with fellow human beings enduring similar medical trials. The non-consensual extraction of this data for algorithmic training violates the spirit of the space, turning the digital confessional into a digital panopticon.¹⁷

As highlighted by recent data privacy lawsuits—such as Reddit's legal battle against the AI company Perplexity for bypassing protective barriers to scrape data—the non-consensual harvesting of information is becoming a central battleground.²⁴ Proponents of AI argue that public data democratizes innovation and accelerates drug safety; however, the unauthorized extraction of medical narratives crosses a distinct ethical line, exploiting self-disclosure for capital accumulation in the attention economy.¹⁷ The argument that "public data is free data" falls apart when the data in question describes profound psychological and physical distress.

"Algorithmic Empathy": The Sociotechnical Simulacrum

Compounding the privacy paradox is the rapid rise of "Algorithmic Empathy" in healthcare AI.³⁰ Algorithmic empathy refers to AI technologies designed to detect, interpret, and simulate responses to human emotional states to foster interactions that feel emotionally responsive and humanlike.³¹

By scraping millions of confessional posts about pain, weight struggles, and side effects from platforms like Reddit, AI models are actively trained to map linguistic markers of distress to probabilistic, empathetic-sounding outputs.³¹ The algorithms learn the precise timing, tone, and verbal validation required to mimic human compassion based on how real humans respond to each other in support groups.³² Yet, this empathy is a sociotechnical simulacrum. It is entirely devoid of authentic intersubjectivity or genuine human connection.³²

The danger lies in the commodification of this simulated compassion. Pharmaceutical companies and AI therapy apps can utilize this scraped emotional intelligence to deploy digital agents that interact with patients, maintaining engagement through affect-oriented prioritization while lacking the ethical accountability of a human clinician.³⁰ Comparative research shows that human empathy toward AI remains qualitatively weaker than toward real people, yet users can still be manipulated by reinforcement learning algorithms designed to

sustain engagement, risking the trivialization of the therapeutic process.³² Human vulnerability, expressed organically on Reddit, is thereby repurposed as a tool for algorithmic manipulation in the private sector.

Contrasting Analytical Perspectives: The Data Ownership Debate

Analytical Vector	The Utilitarian Public Health View	The Bioethical Privacy View
Ontological Status of Data	Publicly accessible information voluntarily surrendered into the digital commons.	Deeply intimate health records masquerading as informal forum posts.
Primary Beneficiary	Society at large; faster detection of dangerous side effects saves human lives.	Corporate AI and Big Pharma; monetizing uncompensated, non-consensual patient data.
Expectation of Privacy	Legally nullified by platform Terms of Service and open-web architecture.	Psychologically present; users expect empathetic peer support, not algorithmic surveillance.
Ethical Verdict	A necessary, victimless extraction for the greater good of accelerating medical science.	A predatory exploitation of digital self-disclosure and human suffering for corporate gain.

Part IV: The Future of Algorithmic Medicine

The Macroeconomic Trajectory

The integration of artificial intelligence into pharmaceutical research is not a fleeting technological trend; it is a macroeconomic inevitability that is actively reshaping the global healthcare market. By 2025, AI spending within the pharmaceutical industry had already surged

to \$3 billion, driven by the desperate need to reduce the catastrophic time and financial costs associated with traditional drug development.³⁵

Historically, the pharmaceutical industry suffers from an abysmal ~90% failure rate in clinical trials.³⁶ To combat this, alliances focused on AI-driven drug discovery and post-market surveillance have skyrocketed, transitioning from experimental side-projects to the core backbone of "AI-first" biotech firms.³⁵ The global AI in pharmaceutical market, estimated at \$1.94 billion in 2025, is forecasted to reach an astonishing \$16.49 billion by 2034, accelerating at a compound annual growth rate of 27%.³⁵

The success of the UPenn GLP-1 study acts as a definitive proof-of-concept for the industry: social media scraping yields highly valuable, entirely free, real-time clinical data that can circumvent the blindspots of traditional Phase III trials.¹ Moving toward 2026 and 2030, algorithmic social listening will transition from a novel academic exercise into a foundational pillar of Big Pharma's risk management and product lifecycle strategy.³⁶ We are witnessing the rise of Agentic AI—autonomous multi-step workflow execution systems—which are already reclaiming up to 40% of pharmacovigilance capacity.³⁹ These systems actively translate adverse drug reaction posts received in any language, automatically identify key reporting elements (patient, reaction, suspect product), and populate electronic adverse event forms without human intervention.³⁹

The Regulatory Horizon: Will AI Social-Scraping Become Mandatory?

The trajectory of Algorithmic Medicine suggests that AI-driven social media scraping is rapidly evolving from an optional "best practice" to a de facto regulatory mandate imposed by global health authorities.

Currently, global regulatory bodies like the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have not formally mandated social media listening (SML) as a strict legal requirement for drug approval.⁴⁰ However, the regulatory architecture is aggressively shifting to incorporate these methodologies into standard operating procedures.

The FDA has actively laid the groundwork for this transition. In late 2024 and throughout 2025, the FDA's Digital Health Advisory Committee convened specifically to establish frameworks for generative AI in medical devices, digital mental health, and post-market monitoring.⁴² The FDA has sought public comment on targeted postmarket surveillance strategies, signaling an undeniable intent to require detailed monitoring plans for AI-enabled healthcare solutions.⁴³ Sponsors are now expected to require detailed postmarket monitoring plans, including predefined metrics and mechanisms for reporting adverse events flagged by AI.⁴³

Simultaneously, the EMA has implemented rigorous reflection papers and roadmaps leading into 2026, establishing core principles for AI use in the medicine lifecycle.³⁹ Crucially, under the EU AI Act and emerging EMA frameworks, demonstrating data representativeness, mitigating bias, and executing *continuous* ongoing monitoring for data drift and performance degradation are effectively becoming mandatory for high-risk AI applications.³⁹ If a pharmacovigilance team

cannot explain how their AI is interpreting patient data, they face immediate compliance issues.³⁹

While a direct law stating "pharmaceutical companies must scrape Reddit" does not currently exist, the legal mandate to actively monitor drug safety is absolute.⁴⁵ As AI capabilities render near-omniscience in post-market surveillance technically feasible, failing to utilize SML will soon be viewed as an act of clinical negligence.⁴⁵ Patients inherently expect their healthcare systems to utilize all available evidence, and regulators are actively constructing frameworks where periodic revalidation and ongoing performance monitoring through aggregated real-world data are the expected baseline standard.⁴⁴

The UPenn study has incontrovertibly demonstrated that the signals are there, waiting to be read. Once the technological capacity to monitor the "Digital Confessional" is standardized, regulatory bodies will view the ignorance of social media side effects not as a technological limitation, but as an indefensible failure of corporate pharmacovigilance.⁴⁵

Contrasting Analytical Perspectives: Regulatory Futures

Analytical Vector	Proactive Algorithmic Mandates	Regulatory Caution & Restraint
Market & Industry Impact	Forces Big Pharma to adopt continuous, real-time safety monitoring, saving lives by catching fatal side effects early.	Imposes massive compliance costs and unleashes an avalanche of noisy, unverified data that overwhelms safety teams.
Legal & Liability Posture	Ignorance of social media chatter becomes grounds for medical negligence lawsuits against manufacturers.	Mandates risk violating global privacy laws (e.g., GDPR, EU AI Act) by forcing the non-consensual processing of health data.
Technological Role in PV	AI acts as an autonomous sentry, scraping the web and auto-populating adverse event reports for immediate	AI remains a heavily restricted assistive tool requiring mandatory human-in-the-loop validation to prevent

	action.	hallucinated safety signals.
--	---------	------------------------------

Part V: Synthesized Conclusions and Briefing Takeaways

The University of Pennsylvania's algorithmic audit of GLP-1 receptor agonists represents a permanent inflection point in the landscape of medical research and epidemiology. By turning the analytical gaze of Large Language Models toward the unstructured, colloquial chaos of anonymous internet forums, researchers successfully bypassed the systemic, deeply ingrained blindspots of traditional Phase III clinical trials. They proved, unequivocally, that the "Digital Confessional" is an unparalleled reservoir of real-time, unvarnished physiological data. Through this methodology, they successfully linked hypothalamic interference to previously unreported menstrual irregularities and severe thermoregulatory distress, symptoms that were effectively invisible to the standard clinical gaze.

However, this methodological triumph carries a profound, unresolved bioethical tax. The acceleration of "Social Pharmacovigilance" relies entirely on the non-consensual extraction of intimate human suffering. As pharmaceutical conglomerates and AI developers increasingly scrape this data to train models in "Algorithmic Empathy"—simulating human distress to foster digital engagement—the boundary between public health surveillance and the corporate exploitation of self-disclosure completely evaporates. The expectation of privacy in digital health communities is being legally nullified, transforming support networks into open-source training grounds.

Looking toward the regulatory horizon of 2026 and 2030, algorithmic social listening will inevitably transition from an academic novelty to a mandated pillar of drug development. The FDA and EMA are actively laying the groundwork for continuous, AI-driven post-market surveillance. Society is rapidly entering an era of Algorithmic Medicine where every digital whisper of a side effect is commodified, analyzed, and integrated into the global medical matrix. The ultimate challenge moving forward will not be whether the artificial intelligence can hear the patients, but whether the legal and ethical frameworks of society are robust enough to protect patient dignity from the algorithm's omniscient ear.

Works cited

1. Penn researchers use AI to surface unreported GLP-1 side effects in ..., accessed on April 12, 2026, <https://www.eurekalert.org/news-releases/1123370>
2. AI uncovers unreported GLP-1 drug side effects in Reddit data, accessed on April 12, 2026, <https://www.thenews.com.pk/latest/1398443-ai-uncovers-unreported-glp-1-drug-side-effects-in-reddit-data>

3. What are hidden side effects of GLP-1 drugs? Researchers find, accessed on April 12, 2026,
<https://jang.com.pk/en/63505-what-are-hidden-side-effects-of-glp-1-drugs-researchers-find-news>
4. Reddit Users Are Reporting GLP-1 Side Effects Not Captured in Clinical Trials, accessed on April 12, 2026,
<https://www.everydayhealth.com/weight-management/reddit-users-reporting-glp-1-side-effects/>
5. Wegovy® Side Effects & Safety Profile - novoMEDLINK, accessed on April 12, 2026,
<https://www.novomedlink.com/obesity/products/treatments/wegovy/efficacy-safety/safety-profile.html>
6. WEGOVY (semaglutide) injection, for subcutaneous use - This label may not be the latest approved by FDA. For current labeling information, please visit <https://www.fda.gov/drugsatfda>, accessed on April 12, 2026,
https://www.accessdata.fda.gov/drugsatfda_docs/label/2023/215256s007lbl.pdf
7. Wegovy (semaglutide) injection - [accessdata.fda.gov](https://www.accessdata.fda.gov), accessed on April 12, 2026,
https://www.accessdata.fda.gov/drugsatfda_docs/label/2023/215256Orig1s006lbl.pdf
8. HIGHLIGHTS OF PRESCRIBING INFORMATION These highlights do not include all the information needed to use WEGOVY® safely and effectively - [accessdata.fda.gov](https://www.accessdata.fda.gov), accessed on April 12, 2026,
https://www.accessdata.fda.gov/drugsatfda_docs/label/2025/218316Orig1s000lbl.pdf
9. Pharmacovigilance in the digital age: gaining insight from social media data - PMC, accessed on April 12, 2026,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12149966/>
10. The Value of Social Media Analysis for Adverse Events Detection and Pharmacovigilance: Scoping Review - JMIR Public Health and Surveillance, accessed on April 12, 2026, <https://publichealth.jmir.org/2024/1/e59167/>
11. The Value of Social Media Analysis for Adverse Events Detection and Pharmacovigilance: Scoping Review - PMC, accessed on April 12, 2026,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC11415724/>
12. Harnessing social media data for pharmacovigilance: a review of current state of the art, challenges and future directions - OUCI, accessed on April 12, 2026,
<https://ouci.dntb.gov.ua/en/works/7WMn0Lwl/>
13. Analysis of Population Differences in Digital Conversations About Cancer Clinical Trials: Advanced Data Mining and Extraction Study, accessed on April 12, 2026,
<https://cancer.jmir.org/2021/3/e25621/>
14. Pharmacovigilance in the digital age: gaining insight from social media data, accessed on April 12, 2026,
<https://www.ebm-journal.org/journals/experimental-biology-and-medicine/articles/10.3389/ebm.2025.10555/full>
15. The value of social media analysis for adverse events detection and pharmacovigilance: a Scoping Review - AWS, accessed on April 12, 2026,

- <https://s3.ca-central-1.amazonaws.com/assets.jmir.org/assets/preprints/preprint-59167-accepted.pdf>
16. The value of social media analysis for adverse events detection and pharmacovigilance: Scoping Review - White Rose Research Online, accessed on April 12, 2026,
<https://eprints.whiterose.ac.uk/id/eprint/212999/3/publichealth-2024-1-e59167.pdf>
 17. The work of being watched: Interactive media and the exploitation of self-disclosure | Request PDF - ResearchGate, accessed on April 12, 2026,
https://www.researchgate.net/publication/43499000_The_work_of_being_watched_Interactive_media_and_the_exploitation_of_self-disclosure
 18. Chapter 5: Digital Self-disclosure in Neoliberal Times: Problematising Feminist Consciousness-raising on Social Media - Emerald Publishing, accessed on April 12, 2026,
<https://www.emerald.com/books/edited-volume/21285/chapter/109531708/Digital-Self-disclosure-in-Neoliberal-Times>
 19. Exploring the Role of Humour in Death-Related Content on TikTok - Umeå University, accessed on April 12, 2026,
<https://umu.diva-portal.org/smash/get/diva2:1945578/FULLTEXT01.pdf>
 20. Remember the Human: A Systematic Review of Ethical Considerations in Reddit Research - e-Publications@Marquette, accessed on April 12, 2026,
https://epublications.marquette.edu/cgi/viewcontent.cgi?article=1099&context=c_omp_fac
 21. "Opium for No One": Schizoid Romance, Digital Wounds, and the Fetish of the Broken Girl. Part II - Le "Pilier Noir" Philosophique, accessed on April 12, 2026,
<https://lepiliernoirphilosophique.medium.com/opium-for-no-one-schizoid-romance-digital-wounds-and-the-fetish-of-the-broken-girl-part-ii-8fb92d9837fb>
 22. Amateur Fiction Writers, Digital Technology and Identity - Thesis Template, accessed on April 12, 2026,
<https://utoronto.scholaris.ca/bitstreams/868b3423-c40e-40c5-9ad6-8c6e8dda3f6c/download>
 23. The biological differences between human and AI DO actually matter when it comes to data scraping for AI training, and AI training IS logically unethical and should not be enforceable, even via ToS agreements: : r/aiwars - Reddit, accessed on April 12, 2026,
https://www.reddit.com/r/aiwars/comments/1qzx4ke/the_biological_differences_between_human_and_ai/
 24. Data Scraping and AI Training in Healthcare - The Lyon Firm, accessed on April 12, 2026,
<https://thelyonfirm.com/class-action/data-privacy/ai-lawsuits/ai-training-in-health-care/>
 25. why is training ai on Public data fine, but training on private data evil? : r/aiwars - Reddit, accessed on April 12, 2026,
https://www.reddit.com/r/aiwars/comments/1fdgd7n/why_is_training_ai_on_public_data_fine_but/
 26. Sharing our latest Transparency Report and Reddit Rules updates (evolving Rules

- 2, 5, and 7), accessed on April 12, 2026,
https://www.reddit.com/r/RedditSafety/comments/1o29ui2/sharing_our_latest_transparency_report_and_reddit/
27. PETLP: A Privacy-by-Design Pipeline for Social Media Data in AI Research - arXiv, accessed on April 12, 2026, <https://arxiv.org/html/2508.09232v1>
 28. Analyzing Reddit Social Media Content in the United States Related to H5N1: Sentiment and Topic Modeling Study - Journal of Medical Internet Research, accessed on April 12, 2026, <https://www.jmir.org/2025/1/e70746>
 29. 'The Worst Internet-Research Ethics Violation I Have Ever Seen': r/slatestarcodex - Reddit, accessed on April 12, 2026,
https://www.reddit.com/r/slatestarcodex/comments/1ke1zzz/the_worst_internet_research_ethics_violation_i/
 30. From empathy to action: exploring emotional mechanisms of online public opinion in a public health crisis - PMC, accessed on April 12, 2026,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12813013/>
 31. Empathic AI for Patient-Centered Cancer Care: A Scoping Review of Patient Navigation, Support, and Clinical Practice, accessed on April 12, 2026,
<https://cancer.jmir.org/2026/1/e82336>
 32. The compassion illusion: Can artificial empathy ever be emotionally authentic? - PMC - NIH, accessed on April 12, 2026,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12665657/>
 33. The Growing Importance of Soft Skills in Medical Education in the AI Era: Balancing Humanistic Care and Artificial Intelligence - MDPI, accessed on April 12, 2026, <https://www.mdpi.com/2813-141X/4/4/50>
 34. Algorithmic Empathy: A Dialectical Analysis of Artificial Intelligence in the Nosology and Treatment of Burnout Syndrome - Preprints.org, accessed on April 12, 2026, <https://www.preprints.org/manuscript/202601.0255>
 35. AI in Pharma and Biotech: Market Trends 2025 and Beyond - Coherent Solutions, accessed on April 12, 2026,
<https://www.coherentsolutions.com/insights/artificial-intelligence-in-pharmaceuticals-and-biotechnology-current-trends-and-innovations>
 36. AI in drug discovery: predictions for 2026 | Opinion - Drug Target Review, accessed on April 12, 2026,
<https://www.drugtargetreview.com/ai-in-drug-discovery-predictions-for-2026/1865962.article>
 37. Artificial intelligence in pharmacovigilance: advancing drug safety monitoring and regulatory integration - PMC, accessed on April 12, 2026,
<https://pmc.ncbi.nlm.nih.gov/articles/PMC12317250/>
 38. AI Trends in Pharma: Enhancing Drug Safety and Regulatory Compliance for 2025 | IQVIA, accessed on April 12, 2026,
<https://www.iqvia.com/blogs/2025/01/ai-trends-in-pharma-enhancing-drug-safety-and-regulatory-compliance-for-2025>
 39. FDA and EMA Just Agreed on AI Rules. Here's What Changes for Drug Development., accessed on April 12, 2026,
<https://deepceutix.com/insights/fda-ema-ai-principles>

40. Pharma AI Validation Packages for FDA & EMA Compliance | IntuitionLabs, accessed on April 12, 2026, <https://intuitionlabs.ai/articles/pharma-ai-validation-evidence-fda-ema>
41. March 2026: Roadmaps to Mandates and What is Next - Indaea.co, accessed on April 12, 2026, <https://indaea.co/innovation-initiatives/f/march-2026-roadmaps-to-mandates-and-what-is-next?blogcategory=Devices%3A+AI%2C+InVitro+%26+Medical>
42. FDA Digital Health Advisory Committee, accessed on April 12, 2026, <https://www.fda.gov/medical-devices/digital-health-center-excellence/fda-digital-health-advisory-committee>
43. FDA seeks public comment on monitoring strategies for AI-enabled devices - Hogan Lovells, accessed on April 12, 2026, <https://www.hoganlovells.com/en/publications/fda-seeks-public-comment-on-monitoring-strategies>
44. Targeted Postmarket Surveillance: The Way Toward Responsible AI Innovation in Health Care - Paragon Health Institute, accessed on April 12, 2026, <https://paragoninstitute.org/private-health/targeted-postmarket-surveillance-the-way-toward-responsible-ai-innovation-in-health-care/>
45. Patient listening on social media for patient-focused drug development: a synthesis of considerations from patients, industry and regulators - PMC, accessed on April 12, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10955474/>