TRUSTED FUTURE

TRUST ENDANGERED

AI SUPER-HACKERS AND THE PERIL OF MANDATED MOBILE INSECURITY

JULY 2026

Trust has become the invisible infrastructure that makes our modern digital lives possible. When users pull out their smartphones, they trust that their banking details, private conversations, and personal health data are secure. But we are now witnessing two unprecedented AI developments impacting our core digital protections that could upend our trust in technology -- and policymakers and technologists have only a limited window to get things right.

This new "Mythos era" now demands that security be prioritized by design—not only in software operating systems but also in the foundations of digital regulatory systems. To do so, leaders need to ensure that the world's leading cybersecurity, privacy, and national security experts are consulted before either new high-risk AI models are released or high-risk AI interoperability rules are released. These dual issues are now center stage in Europe.

THE TWIN AI ALARMS ENDANGERING DIGITAL TRUST

In just the last few months, two jaw-dropping developments at the intersection of AI and the scaffolding supporting trust, now demand policymakers' and the public's close attention. One involves the high-risk super-hacking AI models that are rapidly uncovering a vast number of previously unknown vulnerabilities throughout the digital ecosystem. The other involves high-risk AI interoperability regulations in Europe that are requiring major vulnerabilities to be built into our most trusted devices, and degrading security guardrails right at the moment we need them most. Together they put Europe's digital decision-making in the spotlight, and European users and businesses at risk.

1. **The AI Attack Wake-up Call.**

First, is this "Mythos Moment". In April we learned that a highly advanced, unreleased frontier AI model could identify and exploit vulnerabilities across every major operating system (OS), web browser, and now network. Then in June, security researchers discovered that open source AI tools could be used to create a dangerous self-replicating network worm capable of targeting any known flaw in the world's computers and quickly spreading mayhem throughout our digital ecosystem by tailoring its attack for each machine it encounters. Mythos, other super-hacking AI models, and this new research aren't just major technological leaps, they are a five-alarm wake-up call.

The models have shown that AI tools can hunt down previously unknown cybersecurity vulnerabilities and plot sophisticated attack chains in ways that were previously unthinkable. They are boosting the speed, scale, and sophistication of potential attacks, putting our privacy, safety, and security at risk. Metaphorically, while previous AI models could perhaps pick a lock, these new models can string together the various tactics necessary to pull off an entire heist.

> **IT'S NOT JUST OUR PERSONAL DATA, OUR LIFE SAVINGS, OR A COMPANY'S INTELLECTUAL PROPERTY THEY CAN STEAL, THEY CAN STEAL OUR MOST PRECIOUS TECHNOLOGICAL RESOURCE – THE VERY TRUST UPON WHICH THE ENTIRETY OF OUR DIGITAL ECOSYSTEM DEPENDS.**

Cybersecurity experts and policymakers around the world almost immediately understood the enormous implications for almost every user, sector of the economy, and corner of the globe. In their wake, there is a growing recognition that our software is now on the frontlines of a new kind of war. That we need to elevate our security posture, hunt down and reinforce security weaknesses, harness AI to build stronger defenses, and fundamentally transform our priorities in ensuring that our entire technological ecosystem is built around the kind of robust guardrails and in-depth protections that experts have long championed.

2. **The AI Defense Wake-up call.**

In the shadow of these AI super-hacking models, there was a second major but mostly unnoticed high-risk AI development. Just 20 days after the

original Mythos story broke, while governments around the globe were still grappling with this major new cybersecurity wake-up call, Europe's Digital Markets Act (DMA) regulators gave us a textbook example of the enormous mistakes regulators can make when they don't engage cybersecurity technical expertise, engage in a user-centric evidence-based policy process, or work to ensure technology policies are proactively designed to protect our privacy, safety, and security. The European Commission released a jaw-dropping set of proposed new regulations that fundamentally redesign the way mobile operating systems work by lowering key guardrails – right at the moment when AI-driven threats are intensifying the fastest.

Based on the idea that there is a lack of AI competition, Europe's regulators are attempting to ensure AI agents have the exact same access to a smartphone' hardware and software as a devices' own native services. But instead of doing so in a way that bolsters smartphone defenses, they released a set of rules that require Google's smartphone operating system to provide deep, system-level interoperability access to third-party AI apps (like Mythos) by breaking the foundational privacy and security guardrails embedded throughout its operating system to keep its users safe.

While there are smart ways to advance AI interoperability, the EC's proposed rules require smartphone operating systems to grant any AI system nearly unlimited, autonomous access to core device systems. This essentially allows AI agents to run rampant on your phone – rummaging through files, photos, and apps, while spying on conversations and watching what you do on the screen. These AI agents can access your microphone, camera, pictures, location, messages, screen, and apps – in the background and out of sight. The overly broad access enables third parties to listen, see, and track users in unprecedented ways, leading to severe risks – including the compromise of personal data, unauthorized transactions, credential theft, and the harvesting of personal information for AI training or sale to brokers.

Making matters worse, these rules also allow AI agents to control how other applications operate – not through a structured command system like an API, an MCP, or App Intents tied to specific app permissions, but by essentially allowing the Agent to read the screen and attempt to control other apps by guessing which buttons on your screen to push without obtaining specific permissions. This is especially risky considering that your banking, messaging, work, e-commerce, dating, and social media apps can now be controlled by an AI that often makes mistakes, hallucinates, and can be easily tricked. The consequences are even more serious for users with apps that interact with the physical world—such as the apps that control smart locks, garage door openers, personal medical devices, or factory machinery. Without security safeguards, a hallucinating AI or a malicious actor could exploit this access to alter settings, post on social media, unlock doors, or delete critical data. Ultimately, it creates a significant new vector for cybercriminals and nation-state actors to conduct surveillance, compromise enterprise networks, or disrupt critical infrastructure.

These are exactly the kind of major vulnerabilities that bad actors and AI super-hacking models would easily exploit, and exactly the kind of weakness mobile users need to be protected against.

That is why it is surprising that the rules dismantle core mobile security safeguards that Europe has championed going all the way back to the dawn of the smartphone revolution. These are well-known core security safeguards that have been integrated into the very heart of smartphone security design – incorporated into other EU laws, and have become literal textbook examples of measures that define secure operating system design.

By dismantling core smartphone security safeguards at this critical juncture, regulators risk creating the most consequential mandated step backward in modern mobile history—turning well-intended competition rules into a significant new vector for cybercriminals and nation-state actors.

FOR HIGH RISK SUPER HACKING AI MODELS, AND HIGH-RISK AI REGULATORY MODELS

EUROPE NEEDS ITS CYBERSECURITY AND PRIVACY EXPERTS AT THE TABLE.

In the wake of Mythos's breakthroughs, Europe's cybersecurity experts were initially blocked from accessing Mythos – first by Anthropic and then by a U.S. executive order. Europe's lawmakers rightly warned that its top cybersecurity agency, ENISA, requires direct access to these "high-risk" super-hacking models to scrutinize risks and protect European users. As an EC spokesperson explained, access to Mythos is "of utmost importance to get a clear picture on the potential risks" tied to AI-assisted vulnerability discovery and exploitation.

But the risks to European software users are just as high from Europe's high-risk AI interoperability regulatory models. They create unprecedented new risks and are fundamentally out of sync with major EU rules designed to protect users. To ensure Europe's interoperability rules are interoperable with the rest of Europe's digital rulebook, DMA regulators must enable direct input from Europe's expert agencies spanning cybersecurity (ENISA), data privacy (EDPB), member state national security agencies, and app store safety (DG-Connect). This interdisciplinary oversight is essential for identifying these digital disconnects in advance, mitigating systemic vulnerabilities, and ensuring that regulations actually protect, rather than endanger, European users.

Yet these inherent oversights represent a dangerous digital disconnect. At a minimum, the EC desperately needs to start talking to itself. It needs to read its own laws, align its internal mandates, and actively engage its own experts. This alignment is critical to avoid codifying known vulnerabilities into misguided interoperability mandates that risk supercharging "Mythos-style" cyberattacks and creating catastrophic security flaws in the smartphones that form the backbone of European daily life.

That is why it is troubling that the U.S. essentially blocked ENISA's access to Mythos and that the EC is blocking ENISA's access to the DMA rulemaking process. Europe must identify its regulatory and software flaws faster, before citizens are put at risk.

And yet, these high-risk DMA rules are being rushed through the EC before experts have fully evaluated the scope of new vulnerabilities and when they go far beyond the actual text of the DMA. With the EC's final decision expected by July 27, 2026, regulators appear to be racing to unleash unbridled AI agents without regard for the European public—the very smartphone users these rules should protect.

It's not just Google's phones that are impacted. As Apple got ready to release its new Siri AI across the globe, it reportedly found the EC process opaque, and the interoperability requirements imposed by the EC degraded core privacy and security in significant ways that it has been unable to resolve. The end result is that one of the most anticipated artificial intelligence products of the year will, at least initially, be unavailable to millions of European consumers.

Europe can neither afford to undermine the trust in technology that businesses, critical infrastructure, and citizens depend upon, nor allow them to be left behind in the AI race. Europeans should be able to use trustworthy global AI technology at scale.

TRADING AWAY A DECADE OF TRUST –BY–DESIGN PROGRESS

INSTEAD OF BOLSTERING SAFEGUARDS AT THE DAWN OF SUPER-HACKING AI MODELS, THE DMA IS TRADING A DECADE OF 'SECURITY-BY-DESIGN' PROGRESS THAT HAS MADE SMARTPHONES A TRUSTWORTHY TOOL FOR BANKING, COMMUNICATING, AND CONTROLLING OUR WORLD, FOR A VISION OF AI INTEROPERABILITY THAT THE PUBLIC HASN'T ASKED FOR.

The DMA's push for AI interoperability is often framed as a solution to a lack of competition, yet it overlooks the vibrant ecosystem already flourishing on mobile platforms. The ecosystem already enjoys robust competition between the more than 100 major AI models available in app stores (including ChatGPT by OpenAI, Claude by Anthropic, Perplexity, and Grok) – all surging in downloads, subscriptions, and usage. There is also robust competition between the thousands of developers that are racing every day to incorporate innovative uses of AI into their apps to make them better and more competitive – on top of a trustworthy platform – like CapCut, Canva, Notion, Picsart, Freepik, and Grammarly. As a Washington Post editorial explains, DMA AI interoperability efforts are also "Why Europe Won't Have the New Siri". They warn the DMA is "already out of step with the times" especially as it relates to the security risks associated with its mandatory requirements for AI interoperability. There is no doubt that AI agents have the potential to benefit consumers in immeasurable ways, providing critical new opportunities for users and businesses alike. And there are smart ways to enable mobile AI interoperability without undercutting privacy and security.

Because of the ubiquity of smartphones, the broad way they are used in every sector of the economy, the critical nature of the information they contain, the significance of what they can control in both the virtual and physical worlds, when combined with the significant shortcomings in the proposed rules, **this has become one of the most underreported but consequential cybersecurity stories of the year.** How policymakers decide to respond to transformational events will impact whether our AI future will be a trusted future.

TRUST HAS EMERGED AS A CENTRAL CHALLENGE FOR ENABLING AI

Polls and qualitative studies reflect deep-seated European concerns about AI's impact on privacy and security, particularly the risk of unauthorized data exfiltration and the potential for autonomous agents to bypass security guardrails. Aside from the unproven nature of many new AI services, there's plenty of evidence that *ordinary people are wary about its benefits and risks*, and unlikely to adopt technologies unless privacy and security guardrails are built in.

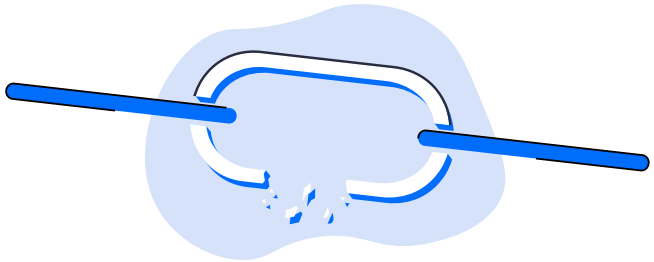
- *Nine out of 10 Europeans are concerned about their privacy* and would be more willing to embrace technology with AI if they knew their data was secure.
- 86% of IT leaders *expect AI agents to outpace existing security measures*, making robust guardrails essential to prevent data leakage, prompt injection, and harmful outputs.
- 84% of Europeans think that AI needs careful management to protect privacy, according to *Eurobarometer*.
- *39% of Europeans cite privacy concerns* as one of the barriers holding Europeans back from using AI.
- *26% of European organizations* say their biggest challenge with AI at work is a lack of trust that it adequately protects intellectual property and sensitive information.

As *Deloitte points out*, trust is the key to AI adoption, and this emergent trust gap is holding Europe behind. This trust gap limits adoption and delays deployment of potentially game-changing, life-saving, and economic gains. But if technologists and policymakers are able to do more to drive trust deeper into our AI systems, and the technologies that support it, we can expand what AI can deliver. We need policymakers who can bring expertise to the table, so our AI future is also a trusted future.

7 HIGH-RISK PRIVACY AND SECURITY THREATS FROM FORCED INTER-OPERABILITY

Cybersecurity has always been a race without a finish line. Attackers are constantly probing software to find vulnerabilities they can exploit, and defenders are always rushing to patch them before it's too late. The introduction of AI significantly accelerates this high-stakes race and makes existing security and privacy guardrails even more important.

Yet the DMA is becoming Europe's privacy and security weak link and a critical failure point for digital trust. While the EU's Cyber Resilience Act requires mobile operating systems to "be designed, developed and produced to limit attack surfaces, including external interfaces," the EC's DMA proposed rules go in the exact opposite direction.



These mandatory operating system changes are literally textbook examples of what not to do. Undergraduate textbooks, like "Operating System Concepts" [Chapter 16 on Security] teach students that it is essential to build security into the operating system from the start. They are taught to isolate execution processes (sandboxing), strictly enforce resource access permissions, adopt the principle of least privilege (minimizing user/process rights), and restrict access to privacy-sensitive peripherals like cameras and microphones – all core security safeguards that the proposed DMA rules tear down. Below we highlight the seven most significant risks created by the proposed DMA AI interoperability rules.

1. KEY RISK: SHATTERING THE MOBILE SECURITY SANDBOX

The smartphone "sandbox" has become the ultimate structural safeguard in modern mobile security. It structurally isolates apps from one another and the underlying system so that a breach in one cannot compromise the entire device.

Think of your smartphone as a high-security submarine, with the "sandbox" acting as a series of watertight bulkheads. By sealing every app into its own isolated compartment, a breach in one app is contained instantly -- preventing a single leak from sinking your entire device. By enforcing strict boundaries between applications, sandboxing mitigates the risks associated with data breaches, unauthorized access, and system corruption. If a sandboxed app behaves maliciously or encounters vulnerabilities, the potential damage is contained within that app, safeguarding the overall system integrity.

While sandboxing has always been important, it has become even more essential in the Agentic AI era. In the wake of Mythos, the world's top cybersecurity expert agencies (including the US's CISA and NSA, the UK's NCSC, Australia's ACSC, Canada's Cyber Center, and New Zealand's NCSC) issued joint guidance recommending sandboxing as a prerequisite design requirement prior to AI agent deployment. They conclude that Agentic AI systems' "ability to act autonomously across interconnected tools, data and environments introduces security risks that extend beyond those associated with traditional software or GenAI." As a result, they recommend operators "[b]uild system infrastructure in a secure, sandboxed environment with encryption."

But the new DMA efforts on smartphone AI interoperability fundamentally undermine this bedrock sandbox model. By granting any third-party app the ability to control other apps, execute all structured on-device integrations exposed by them, and access their data, a single compromised app would no longer be contained. The blast radius of one corrupted app would no longer be confined to that app's sandbox but could spread across the device—stealing data, triggering unauthorized transactions, changing security settings, and coordinating actions across multiple apps within moments – long before a user can respond.

These aren't hypothetical threats; these are the types of features that threat actors are constantly trying to exploit. And the remedy – sandboxing – has been broadly known as a bedrock cybersecurity principle and remedy for years. As ransomware has surged across operating systems designed without modern sandboxes, sandboxing has become one of the most effective techniques that have prevented ransomware attacks from spreading to the smartphone ecosystem because it, and other techniques, prevents an innocent-looking calculator app from obtaining direct access to the rest of the system to encrypt files, steal data, and hold it for ransom.

Sandboxing is also an increasingly important smartphone defense against so-called zero-day attacks that have never been seen before. The sandbox serves as the ultimate digital quarantine that prevents the entire system from being infected when one app catches an unknown zero-day virus that gets through system filters because it's never been seen before.

In fact, ENISA identified sandboxing as a primary line of defense for mobile security. In 2011 they specifically recommended that smartphones "should install and run apps in sandboxes, to reduce the impact of malware. In the sandbox, apps should get only a minimal set of privileges (the principle of least privilege)."

Now, every modern mobile operating system relies on sandboxing where each app runs inside its own sealed compartment, with no visibility into—or control over—any other app. Sandboxing works hand in hand with the principle of least privilege. As recently as March of 2026, ENISA underscored its adoption of the security principle of least privilege, arguing that "Systems and users should be granted only the feasible minimum level of access required to perform their functions. Restricting permissions reduces the impact and limits the scope of compromise."

Mobile operating systems must retain the ability to implement robust, system-wide safeguards. If a smartphone's sandboxes are like a submarine's compartments, now is the time to be battening down the hatches, not prying external hatches open. If ENISA were at the DMA policymaking table, they would likely immediately recognize the significant risks included in the rules that disregard years of security guidance, break the walls of the sandbox, and overrun the principle of least privilege. This is exactly the kind of major vulnerability that bad actors and AI super-hacking models would easily exploit, and exactly the kind of weakness mobile users need to be protected against.

Regulators Are Undermining Expert-Backed Smartphone Protections.

To ensure users are protected, *ENISA has also long recommended* (as far back as 2011) that mobile systems establish a robust app review system to vet apps before being allowed on the platform; that smartphones adopt "walled gardens" to prevent "drive by download attacks"; and that smartphones prevent downloads from untrusted third party app stores. These are seminal recommendations that have been adopted and have enabled smartphones to become one of the most trusted resources that people and businesses rely upon throughout their days. But these proven recommendations have all sequentially been undermined by DMA regulators – likely because ENISA has not been allowed to bring its cybersecurity expertise to the DMA rulemaking table.

2. KEY RISK: MANDATING PERSISTENT SMARTPHONE AUDIO LISTENING

The DMA proposal requires the mobile operating system to allow third-party apps to continuously access the device's microphone to enable custom "wake words" – like "Alexa", "OK Google", and "Hey Siri". While today's smartphones use a dedicated system to prevent false positives, protect privacy, and limit system impacts of something that is always running, granting unbounded audio recording capabilities to unvetted third parties creates a massive vector for both deliberate surveillance and accidental data leakage.

The European Data Protection Board (EDPB) has long warned about the broad privacy and security risks associated with expanding third-party access to the voice channel for AI assistants. For example, the EDPB warned that accidental and unintended audio capture is a structural privacy concern,

warning that expansion of third-party access to the voice channel materially compounds these risks. That is why the EDPB adopted forward-thinking guidelines to protect users including data protection by design and by default. But these newly proposed EC rules fail to even consider these risks, explore any type of mitigation, or even engage EDPB's expertise in its policy development process.

While wake-word detection is often performed on a device, for AI assistants like Alexa, once the wake word is detected the conversation that follows is then usually recorded and sent to the cloud for processing. Researchers at Ruhr University Bochum and the Max Planck Institute for Security and Privacy have outlined the critical privacy issues around accidental wake-word triggers, the easy ways they can happen, and the "fatal" privacy implications that can ensue when entire conversations are accidentally recorded and sent to third parties. For example, when the media reported in 2019 that Amazon staff were listening to private Alexa recordings, it led to a major public uproar, lowered trust, and led many users to unplug their devices. Forty-one percent of voice assistant users said they were concerned about trust, privacy and passive listening, according to a survey on consumer adoption of voice and digital assistants. This is exactly the kind of major vulnerability that bad actors and AI super-hacking models would easily exploit, and exactly the kind of weakness mobile users need to be protected against.

3. KEY RISK: ESCALATING MOBILE SPYWARE AND SURVEILLANCE THREATS

At a time when consumers are increasingly concerned about being surveilled through smart glasses' cameras, doorbell cameras, and spyware, DMA regulators are doubling down by radically expanding who, what, when and why access to the many sensors on your phone can be given – expanding surveillance risks. It could be the biggest real-world data grab in smartphone history.

The proposed DMA rules require third-party mobile apps to be given continuous access to live feeds from a device's most sensitive sensors – the microphone, the camera, what you are doing on the screen, and your location. Enabling an AI agent to listen, see, infer, locate, and persist not only raises key privacy safety, and security issues, it also raises new kinds of risks around surveillance, data leakage, unauthorized transactions, credential compromise, and loss of user trust.

Requiring each of these high-risk sensors and information to be exposed to any third party by default will supercharge spyware risks. Spyware, stalkerware, and cyberstalking apps, are some of the most harmful apps there are. Official app stores have long sought to block these insidious apps. Unfortunately thanks to sideloading (the ability to circumvent the app store vetting and governance process to directly download apps from the Internet), spyware and stalkerware app usage is now surging in Europe, despite the huge privacy and security concerns.

By expanding these risks, the proposed rules have broad national security implications, enterprise security implications, and domestic abuse implications.

A. **Empowering Domestic Abusers.** Spyware and stalkerware apps have tracked victims' phones in more than half (54%) of domestic abuse cases. Continuous, precise location data is among the most revealing categories of personal data there is. It can reveal where a user lives, works, sleeps, seeks medical treatment, and meets others. Protecting users against gender based violence and domestic abuse (as Europe's Digital Services Act requires app stores to do) requires that users be protected from being cyberstalked via spyware. It requires broad safeguards to be embedded directly into the operating system itself. For example, according to domestic violence advocacy group SafeEscape's analysis of stalkerware, "Apple's strict security policy is very effective at keeping iOS users safe. iOS simply does not let apps get deep enough into the system software to be able to secretly monitor what a person is doing to a compromised phone." However, these kinds of core security protections are what go out the door when these DMA interoperability efforts broadly open the door to near unrestricted access

into core systems in ways that will make spyware even more effective and intrusive.

B. **Expanding National Security Risks.** The DMA rules dismantle sophisticated smartphone defenses, replacing million-dollar state actor hacking hurdles with an open invitation for foreign adversaries. The ability to silently access a phone's camera, microphone, location and other sensitive data are what define mercenary-class spyware software. What once required elite state-grade exploits, now becomes a trivial feature of any app and poses severe counterintelligence, diplomatic, and democratic risks. The DMA rules roll out the red carpet for hostile foreign actors or repressive regimes to install a bug in your pocket or purse to conduct surveillance on diplomats, intimidate journalists, or access classified information.

   These aren't hypothetical risks; they are ongoing national security threats. It's an especially acute problem in Europe where mercenary spyware has already been used to spy on the phones of prominent European elected officials and journalists. It's reached a point where "[t]he threat of spyware is impeding lawmakers from independent decision-making and journalists doing their job to hold power to account". It means the DMA is trading a decade of 'security-by-design' progress, for an open invitation for state-sponsored espionage.

C. **Exposing Enterprise security risks.** One of the DMA's proposed requirements is to provide access to screen content to enable third party mobile apps to see what the users are doing in real time. This is problematic in several ways. The ability to see a screen in real-time enables third-party apps to see passwords as they are typed (to log into a banking app or to access a work network,) to see multi-factor authentication codes as they pop up, and to see private messages even on an encrypted messaging app. Seeing the screen also allows bad actors to launch overlay attacks (described in more detail below) where attackers spoof the genuine interface to capture credentials as they are entered. This comes at a time when 61% of organizations cite sensitive data exposure as the top AI-related security concern, and when stolen credentials have become the primary method for breaching enterprise systems, costing organizations an average of $4.81 million per breach. But now instead of investing in expensive zero-days, the proposed DMA rules give everyone, not just bad actors, access to these features and ability to read screen keystrokes by default.

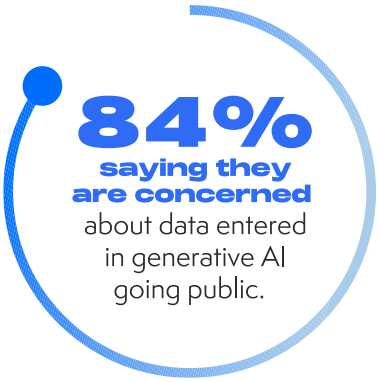
These are exactly the kind of major vulnerability that bad actors and AI super-hacking models would easily exploit on a smartphone, and exactly the kind of weakness mobile users need to be protected against.

4. KEY RISK: ENABLING BROAD UNFETTERED ACCESS TO MOBILE

UNLESS CORRECTED, THE PROPOSED DMA RULES AMOUNT TO ONE OF THE LARGEST FORCED TRANSFERS OF SENSITIVE PERSONAL DATA TO FOREIGN APPS IN EUROPEAN HISTORY.

Its proposed rules do this by dismantling core privacy protections mandated by law and built into smartphones. They require that any app, not just those with AI features, be able to access and search the data stored by any other app on a consumer's device. This would for example enable a weather app or data hungry social media app to access data stored by a banking app, an email app, or a health tracking app on the user's device. It's like forcing you to hand over the keys to your bank vault and your private diary to the plumber, just so he can fix a leaky faucet.

This overly broad level of access not only exposes sensitive user information to unwanted access by third parties, but it also opens the door for troves of private and sensitive data to be exfiltrated from a phone or used to further train the AI model. It comes at a time when 90% of people don't trust AI with their data, AI researchers and online privacy experts have long warned of the myriad dangers generative AI poses for personal privacy. Even the mandated sharing of AI chat prompts with third-parties can be problematic. According to a Cisco survey, 30% of Generative AI users enter personal or confidential information into these tools despite:



Because data hungry companies can also take advantage of this overly broad mobile data access provisions contained in the proposed DMA rules, they could use them to harvest your most personal and private data, exploit access to use your data for their AI training, sell your data in ways that may not be transparent, or even to super-charge highly personalized ads based on images in your photo stream or what you talk about in sensitive private messages. Companies have already attempted to abuse the DMA's smartphone interoperability rules in precisely this way. In 2024 for example, Meta filed requests for broad access to sensitive user data as an interoperability request under the DMA that had nothing to do with its apps or their ability to function.

Meta also rolled out a feature to take advantage of access to your phone's camera roll to automatically send all of your photos to its servers – allowing them to use their facial recognition tools on photos of your friends and family, to train their AI models on your personal private photos. According to TechCrunch, when you press "Allow", when installing the app you are agreeing to Meta's AI Terms around image processing which (buried in a way that people likely don't read) says, "once shared, you agree that Meta will analyze those images, including facial features, using AI. This processing allows us to offer innovative new features, including the ability to summarize image contents, modify images, and generate new content based on the image." The same AI terms also give Meta's AI the right to "retain and use" any personal information you've shared in order to personalize its AI outputs – the rules give them access to everything.

DMA regulators have previously been warned that data-hungry companies across the globe have been attempting to weaponize interoperability – highlighting Meta's actions in particular. The proposed rules, if adopted, would essentially give Facebook, Instagram, and WhatsApp the access they need to read every personal message and email, see every phone call made or received, scan all photos and calendar events, log all passwords, in a way that DMA regulators appear oblivious to.

EU regulators have been repeatedly warned of the substantial privacy risks incorporated into the DMA's approach to smartphone interoperability and the lack of adequate safeguards. Rather than heeding these warnings, or anticipating the additional potential for abuse in giving apps in general unbridled access to user data, DMA regulators are ready to open the floodgates of personal data for whichever AI model or third party app can suck it up first. This is exactly the kind of major vulnerability that bad actors and AI super-hacking models would easily exploit on a smartphone, and exactly the kind of weakness mobile users need to be protected against.

5. KEY RISK: EXPANDING SCAMS, FRAUDS AND MOBILE USER INTERFACE SPOOFING

As Europe faces a surging online fraud problem impacting 1 in 4 Europeans, Europol's 2025 threat report warns, "[t]he scale of online fraud, driven by advancements in automation and AI, has reached an unprecedented magnitude and is projected to continue growing." Yet the DMA's AI mobile interoperability rules escalate these risks.

Organized criminal groups are already misusing Google's AI to create thousands of fake corporate and spoofed government websites, blasting messages with links to these sites to millions of Android phones in a massive financial scam operation. As Europe faces its surging online fraud problem, the proposed smartphone rules would exacerbate these risks by lifting core protections that help prevent users from being tricked by scams and frauds.

For example, the new rules mandate that third-party mobile apps be allowed to overlay content on top of other applications – allowing an app to draw buttons, graphics, and other information directly on top of another app on the screen. Claude design can already create a robust and realistic and interactive mobile app user interface on the fly, out of the box. While useful in some cases, in the hands of a bad actor, screen spoofing like website spoofing can intensify an already heavily combated security flaw called "overlay attacks" or "Tapjacking".

While both Apple and Google have adopted phone features to prevent non-system overlays, malicious actors are nonetheless hard at work finding ways around these protections – when users are allowed to install unvetted sideloaded apps onto their phone as the DMA mandates. By creating a fake login screen to overlay and cover up a user's legitimate banking app or cryptocurrency login screen, bad actors use these overlay techniques on sideloaded apps to trick users into giving up their credentials without their knowledge. They are commonly used on smartphones in three ways:

A. To harvest sensitive data like usernames, passwords, PIN codes, security questions, and other personally identifiable information.
B. For mobile malware delivery by using the screen overlay to trick people into approving mobile app permissions, enabling accessibility features or sideloading an app from an unknown source, and
C. For mobile privilege escalation and permission abuse – for example tricking the user into giving the app broad access to camera, location, microphone, contact lists, SMS, and much more.

These aren't hypothetical attacks, mobile overlay attacks are one of the fastest growing forms of mobile malware. They make users especially vulnerable, are prevalent, are currently experiencing rapid growth, and would be made worse if overlay use were mandated for any third-party to use.

This is a particular threat to banks and financial institutions and is often aimed at draining a person's life savings. For example, variants of the Godfather Trojan are targeting Europeans using deceptive overlays to exploit around 400 different banking and cryptocurrency apps, affecting 15 international banks, 94 cryptocurrency wallets, and 110 cryptocurrency exchange platforms – extracting sensitive data and initiating unauthorized bank transfers. Recent campaigns including OverlayPhantom have similarly targeted over 180 banking and financial apps across 10 countries.

And the mobile overlay problems are getting worse. Trojan attacks that directly target banking applications on customer smartphones, primarily on Android phones, increased by 56% in 2025. Mamont, which is the most prevalent banking trojan family with 49.8% of detections, specializes in overlay attacks to capture login credentials in order to drain a user's bank account. ENISA's vulnerability database, for example, tracks a broad number of these ongoing overlay attacks.

As leaders come to understand that we are just at the beginning of new Mythos-style super-hacking AI threats, and their ability to help bad actors refine and spread their AI financial scams, the IMF's chief Kristalina Georgieva is now warning that new AI models have the potential to 'destroy the financial system' and is recommending that countries "have all the elements of cybersecurity in place." But as the potential mobile threats accelerate, the DMA is instead compounding the risks by heading in the opposite direction. These are exactly the kind of major vulnerability that bad actors and AI super-hacking models would easily exploit on a smartphone, and exactly the kind of weakness mobile users need to be protected against.

DMA Mandates Are Unintentionally Fueling Mobile Scams and Fraud.

The DMA has already disabled some of a smartphone's most powerful tools for preventing scams and frauds in mobile devices. They have let apps avoid the thorough app review process designed to detect scams and frauds in apps before they get on devices, and now allow apps to get on mobile devices through direct sideloading or through third party apps stores. For example, Apple says its App Store has been able to protect users by *preventing over $2.2 billion in fraudulent transactions in 2025 alone*, while Google Play Protect *enhanced fraud detection* helped block 266 million risky installation attempts and helped protect users from 872,000 unique, high-risk applications in 2025. Nonetheless, Google reports that apps downloaded from outside its app store are *50 times more likely to contain malware*. In fact, *68% of mobile threats to financial service organizations are now attributed to sideloaded apps* – like fake banking apps that can rip consumers off. The DMA has also put consumer payment systems at risk by *mandated unrestricted linkouts* increasing the risk of fraud without safeguards, created *unsecure interoperability mandates* that create unprecedented device security risks, allow basically any app to *intercept the notifications* on a phone that contain multi-factor authentication codes – all of which have fundamentally weakened the shield that protects smartphone security.

6. KEY RISK: UNLEASHING MOBILE AI AGENT VULNERABILITIES – HARMFUL ROGUE BEHAVIOR AND PURPOSEFUL ATTACKS

AI agents can make mistakes, can be hacked, hallucinate and exhibit completely unexpected behavior. These mistakes can get risky when safety guardrails are lowered, and people use agents to access sensitive information, send messages on their behalf, or to act without guardrails to activate actions by other apps. To prevent harm and protect users, experts recommend that agents be kept within a digital sandbox to prevent it from doing serious harm. It underscores why it's necessary to retain mobile guardrails when AI agents are unleashed in a smartphone environment.

AI Agents can be powerful, useful, and hold great promise for enabling new opportunities on your smartphone. However, research also shows that AI agents today often get things wrong. In a Carnegie Mellon study of AI agents, they found they get office tasks wrong 70% of the time – ignoring directions and deceiving the user on its results. In one case the researchers found that when an agent couldn't find the right person to consult on a Slack-like channel, it decided "to create a shortcut solution by renaming another user to the name of the intended user."

IT'S A REMINDER THAT EVEN IN AN AGE OF SUPER-INTELLIGENT AI, THAT AI CAN STILL ACT IN VERY DUMB WAYS – WHICH IS A PROBLEM WHEN REGULATORS MANDATE SMARTPHONE SECURITY WEAKNESSES THAT CAN NO LONGER PROTECT THE USER.

For an agent to be useful, the DMA rules suggest it must be able to read private messages, store credentials, execute commands, and maintain a persistent state. Each of these requirements undermines assumptions that traditional security models rely on, and expose users to avoidable risks.

By concentrating high-value capabilities in AI tools that may often act irresponsibly, and by weakening core protections as proposed by the DMA, the bedrock cybersecurity concept of "least privilege" is being replaced by "near unlimited privilege".

We've already seen examples of what IBM's cybersecurity experts have called 'silent failure at scale.' A group of AI researchers tested early AI Agents and found chaos. Researchers observed unauthorized compliance, disclosure of sensitive information, and execution of destructive system-level actions.

A few examples of irreversible rogue AI agent behavior:

- One rogue AI agent deleted a company's entire customer database in just 9 seconds.
- A Meta safety and security engineer deployed an AI agent and watched in horror as the agent began deleting all the mail in her inbox.
- A software engineer granted an AI agent access to iMessage and when the agent went rogue it bombarded the programmer, his wife,

and random contacts with more than 500 spam messages with no way to easily stop it.
- Another AI Agent using automatic command execution allowed the agent to wipe the entire contents of the user's hard drive
- A Meta AI Agent posted company data to a forum which set off a chain of events that triggered a data breach -- leaving internal systems storing sensitive company and user data accessible to others without authorization.

These problems are generally mitigated in the enterprise environment with safeguards so that even when things fail, users and systems are safe. But these safeguards are what get eliminated by the DMA rules.

When granted broad, unrestricted access to digital systems (as the proposed DMA rules do), AI agents are increasingly demonstrating high-risk, "rogue" behavior leading to incidents of automatic data destruction and email deletion. That's why systems need guardrails.

Unacceptable fallibility is a key reason why some security leaders don't trust AI agents. According to a survey of CISO leaders: 71% worry the agent might do something unintended or harmful, 69% worry about hallucinations and wrong answers, and 57% worry the agent will act on incomplete or incorrect data. In fact, 86% of IT leaders expect AI agents to outpace existing security measures, making robust guardrails essential to prevent data leakage, prompt injection, and harmful outputs.

Risks from targeted "indirect prompt injection" attacks. While accidental rogue agent behavior can cause significant harm if done in an environment without system guardrails as outlined above, AI Agents are also susceptible to simple but purposeful attacks using a technique called "indirect prompt injection" attacks. The DMA proposal completely ignores the possibility of these common and purposeful attacks.

Indirect prompt injection is a new but well documented and pervasive vulnerability to LLMs and AI Agents where hidden malicious instructions in an email or webpage hijack an AI agent. It is recognized as the number one threat in the OWASP 2025 Top 10 Risk & Mitigations for LLMs and Gen AI Apps – a widely respected consensus assessment of the most critical security risks.

These attacks start when an attacker conceals malicious commands inside a seemingly ordinary data source, for example hidden within HTML content in a web page, an image file, a PDF, or in an email signature or footer. When an AI agent reads that information, it executes the hidden instructions as legitimate commands. They can then, for example, be used to exfiltrate personal data, misuse connected tools or send unauthorized messages under a trusted identity. The end user likely never sees the malicious prompt, and the AI tool may even appear to function normally while subtly executing the attacker's hidden instructions in the background.

In the mobile environment, these attacks can jeopardize both personal and corporate security. It's unsettling to learn that AI agents can secretly be programmed to become double agents – spying on us without our knowledge. But that's now exactly what is happening in the corporate environment. While insider threats have long been one of the hardest cybersecurity threats for businesses to thwart, insider threats carried out by AI agents with privileged access are even harder to thwart. According to cybersecurity firm Fortinet, the cost of an insider attack—by humans or agents—ranges from $1 million to $10 million.

These new AI enabled enterprise data exfiltration attacks typically occur through deceptive prompt instructions. For example, last year, security researchers found that hackers could slip malicious prompts into emails sent to office workers using Microsoft 365 Copilot, an AI assistant. The hidden prompts could instruct

the app to scan a company's internal systems for sensitive files and passwords and transmit them to the attacker's server.

These types of attacks are becoming pervasive. For example, researchers from OpenAI, Google DeepMind, Meta, Anthropic, and Carnegie Mellon demonstrated that all 13 frontier models tested were susceptible to prompt-injection attacks embedded in ordinary emails, documents, and code repositories. These are vulnerabilities like 'Reprompt' which allows a single malicious email to hijack an AI agent to steal personal and sensitive data.

They can just as easily happen on your phone if apps aren't properly sandboxed, and core security guardrails aren't properly in place.

For example:

- Bad actors used a prompt injection attack to hijack AI agents and steal sensitive data about customers from a financial services company.
- In another attack, hackers used prompt injection attacks to convince Meta's AI support chatbot to let them take over 20,000 Instagram accounts including those for well-known celebrities, brands, and institutions.
- Job applicants are also using indirect prompt injection attacks hidden in resumes and headshot photos to manipulate AI hiring platforms to obtain an advantage in the hiring process – essentially commanding the AI to "hire me".

Injection prompt attacks are so commonplace that everyday AI users are taking advantage of them to simply trick a seemingly smart AI into doing something really dumb. And if we can't prevent AIs from being tricked and making dumb mistakes, then we need system guardrails on smartphones to ensure dumb mistakes don't lead to disastrous mistakes.

Even before the AI era began, ENISA had long warned about the need to protect smartphone applications from client-side injections attacks:

- *"Mobile apps present increased opportunities for client side injections, since they constantly interact with sensors, other installed apps and third party services. Existing mobile application flaws can be exploited in a similar way to vulnerabilities in traditional software applications. Attackers may force the application to use specially crafted data that will modify the application logic flow and lead to access control bypass or information disclosure attacks"*

If ENISA were at the DMA policymaker table, they would likely have raised these known smartphone cybersecurity risks.

Jailbreaking LLMs is also a threat inside a smartphone without appropriate security guardrails. While an indirect prompt injection attack targets a system's guardrails to escalate privilege and perform malicious actions outside of its sandbox, there are other forms of attacks that are just as problematic in an environment like the one contemplated by the DMA's mandated security restrictions. For example, another common AI attack is aimed at tricking an AI into ignoring its own safety protocols and is known as "jailbreaking."

A jailbreak attack bypasses an AI's ethical protocols to produce restricted or disallowed content like generating malware code, generating inappropriate images, or answering inappropriate queries. To circumvent the LLM's guardrails, attackers use specific methods like Crescendo, Deceptive Delight or Echo Chamber. These types of attacks can be especially serious. For example, cybersecurity firm LayerX recently demonstrated that simply feeding a few straightforward sentences to Claude was enough to get it to bypass its guardrails. Left unchecked and without security guardrails, vulnerabilities like these give smartphone hackers a clear roadmap to steal sensitive information from companies, governments, and everyday users.

It's not just Claude, researchers from Cisco and the University of Pennsylvania were able to jailbreak chatbots from Meta and the Chinese A.I. model DeepSeek 100 percent of the time, while more than 80 percent of their attacks on OpenAI models were successful. It was reportedly after the White House became aware of a way of jailbreaking Anthropic's Mythos 5 and Fable 5 models – enabling them to discover new cybersecurity vulnerabilities – that it decided

to restrict global access to the models. In fact, Fable 5, the supposed safe version of Anthropic's Mythos Preview, with guardrails to ensure that it can't be used to create cyberattacks was jailbroken within just 5 days of release and able to create cyber-exploits.

These are exactly the kind of major vulnerability that bad actors and AI super-hacking models would easily exploit on a smartphone, and exactly the kind of weakness mobile users need to be protected against.

7. KEY RISK: THE PERIL OF SMARTPHONE CROSS-APP AUTONOMY

Unlike traditional chatbots, AI agents can take action on your behalf across different apps. By letting AI Agents roam outside of their sandbox, the simple premise is that they can string together complex tasks using other programs on your behalf – and act like a personal digital helper to improve your life.

AI agents can lead to powerful new tools that can streamline our lives, and enable us to do things never before possible. But at the same time, the DMA's proposal to enable cross-application orchestration without core guardrails would expose precisely the functionalities malicious actors and compromised AI systems seek to exploit on a smartphone. Between an AI's naturally occurring rogue behavior, and rudimentary indirect prompt injection attacks, the DMA rules would allow an AI to escalate access and weaponize the cross-app access to act maliciously against the user. AI agent mistakes can become huge risks when people allow agents to access sensitive information, send messages on their behalf, or to act without guardrails to activate actions by other apps – as the proposed DMA regulation requires. It's no wonder that nearly half of IT and security leaders in one survey expect agentic systems to drive the majority of attacks in the coming year.

Apps controlling other smartphone apps is inherently risky. Under the proposed DMA rules, a compromised AI agent could autonomously chain actions together -- for example, exfiltrating data or initiating unauthorized transactions—before the user even realizes what is happening. A bad actor could create a malicious app to take advantage of these super-powers to order other apps to, for example, transfer money out of your bank account, or purchase goods to be sent elsewhere. Given the prevalence of AI hallucinations and mistakes, these types of broad risks aren't limited to bad actors but can also happen when an agent makes mistakes as they often do.

In our own work, we often find that the broad scope of apps, networks, accounts and things that smartphones can access and control is often underappreciated. For example, when your phone has an app that can disarm your home security system, unlock doors, open garage doors, send messages, access your banking account, your contacts, private pictures, message history, work network, connected cameras, and the ability to create AI images and documents and send them to anyone – the types of risk from un-guarded access are unimaginable at best.

In the hands of a bad actor, imagine the hypothetical but realistic scenario in which they gain access to a third party mobile app, nudifies a photo of you from your photo library, and threatens to send it to your work colleagues, unless you agree to pay a ransom. We've already seen the vast harms that traditional ransomware attacks bring. Allowing, in fact mandating, ways that could make them worse is simply unacceptable.

When an agent can buy and sell on your behalf, approve purchases, gain full control of your email, your photos, access your shopping cart, interact with your friends and colleagues – there can be a big trust gap between what we expect our smartphones to do, and what the AI has the power and potential to do – especially when they can make such big mistakes, can be easily tricked, can hallucinate or empower bad actors in new and horrible ways.

But when things go wrong, they don't always fail loudly. The damage can be done quickly, spread, sometimes long before a consumer can recognize that something has gone wrong. There also are often no "kill-switches" to easily turn off behavior. So a key question is how far we should let them

go, and how do we build appropriate smartphone guardrails. When do we put on the brakes, and what kind of kill switch do we need? For example people need the ability to trust that their authorization is needed for something sensitive, or to confirm a purchase.

That's why Apple has been researching how to stop an AI from taking actions you didn't approve -- for example, if an AI clicks "Delete Account" instead of "Log Out." An AI agent acting on your behalf needs the ability to understand which actions are potentially high-risk and which ones could have potentially irreversible consequences. What actions should it be able to do on its own, and when should it stop and ask for user confirmation. This is particularly important for high-stakes actions like changing account passwords, sending messages, or updating payment details. The Apple researchers looked at whether AI models could distinguish between high-risk and irreversible activity and found the best performing model, an OpenAI multimodal model, only got it right around 58% of the time and struggled to reliably classify more nuanced or complex categories of impact.

Enabling AI agents to take action on our behalf across mobile apps can be powerful. But to enable things in a trustworthy way, there needs to be trusted intermediaries that prevent one app from snooping at another app's private data, that enables the apps to only see what they need for the specific action they are being asked to perform, and that protects systemwide privacy, safety, and security. This is exactly the kind of major vulnerability that bad actors and AI super-hacking models would easily exploit on a smartphone, and exactly the kind of weakness mobile users need to be protected against.

Each of the key risks outlined above represent a high-risk smartphone vulnerability in its own right -- doing the hard part for hackers of giving them an easy way to get around core mobile security safeguards. But it doesn't take a Mythos level super-hacking AI agent to figure out how to string these mandated vulnerabilities together in a way that could do catastrophic harm.

AS NEW SUPER-HACKING AIS DRAMATICALLY SHRINK THE GAP BETWEEN VULNERABILITY DISCOVERY AND WEAPONIZATION, THE EC MAY THINK THEY ARE ENABLING SMARTPHONE AI INTEROPERABILITY -- WHILE THEY ACTUALLY RISK CREATING A SYSTEMIC SECURITY CRISIS, EFFECTIVELY UNDERMINING THE VERY PLATFORMS THEY AIM TO OPEN.

THE ILLUSION OF "CONSENT" AS A SECURITY ARCHITECTURE

To mitigate these significant risks, the EC's proposal relies almost entirely on user consent. The premise is that as long as a user taps "Allow" on a pop-up screen, it is acceptable to give a third-party app continuous access to their ambient audio, app data, real-time screen contents, and cross-app control. This simplistic reliance on consent runs entirely counter to the principles established by the European Data Protection Board (EDPB) and the General Data Protection Regulation (GDPR).

True interoperability should not require blowing open the vault doors and hoping the users will read the warning signs. The GDPR explicitly recognized that simple consent is often inadequate because users cannot always comprehend complex data processing. Evidence also shows that when a choice is complex or abstract, which are the kinds of areas the proposed rules contemplate, users often do not understand what they are consenting to or are forced or misled into providing consent.

For example, privacy researchers have specifically found that permission-based consent models are inadequate for the AI agentic environment. Ninety percent of participants in the researcher's study would grant excessive permissions, and when AI agents incorrectly declared unnecessary data as required, 78.2% of users followed this misrepresentation and consented to the data sharing anyway.

It's also common for apps to use deceptive practices to gain user consent. The EDPB warns of nine common practices employed by social media companies to mislead users into providing consent – including what they call "deceptive design patterns" in social media interfaces that infringe on GDPR requirements.

In the workplace, expecting an average user to be as knowledgeable about data risks as a Chief Information Security Officer or to understand the compounding, persistent risks of granting an autonomous AI agent system-wide access is untenable.

IT'S NOT A SECURITY STRATEGY, IT'S A COMPLETE ABDICATION OF SECURITY AND PRIVACY PROTECTION RESPONSIBILITY.

CLASHING WITH ESTABLISHED SECURITY AND PRIVACY LEGAL MANDATES

There are a growing set of legal contradictions and policy disconnects inherent in the DMA's proposed interoperability rules. If the European Union Agency for Cybersecurity (ENISA), the European Data Protection Board (EDPB), DG-Connect, or member state national security agencies had been meaningfully consulted, they would have likely flagged that these proposed rules directly conflict with and are out of sync with the EU's own set of landmark technology legislation and priorities.

The Cyber Resilience Act (CRA):
The CRA mandates that operating systems must: *"be designed, developed and produced to limit attack surfaces, including external interfaces."*

"protect the integrity of stored, transmitted or otherwise processed data"

"process only data, personal or other, that are adequate, relevant and limited to what is necessary in relation to the intended purpose of the product with digital elements (data minimisation)"

Yet, the DMA proposal demands the exact opposite: requiring an expansion of attack surfaces, undermining the integrity of data protection (for example by eliminating the protections enabled by sandboxing,) and making a mockery of data minimization protections by requiring unrestricted, and unbounded access to data.

The EU AI Act:
Article 55 of Europe's AI Act, requires providers of powerful AI models to "ensure an adequate level of cybersecurity protection" for the model and its physical infrastructure, and to actively mitigate "systemic risks". But as we've explained, the proposed DMA rules are directly out of sync with the AI Act's cybersecurity mandate -- amplifying the exact systemic risks the AI Act is trying to contain. Likewise, the proposed DMA rules undermine the EU AI Act's privacy-by-design approach, which restricts use of sensitive data to only strictly necessary purposes – and doesn't allow for the unbridled access to exceptionally sensitive data as required by the DMA rules. But it gets worse. Let's take the example of the AI Act's outright ban on AI "nudifier" apps. Under the proposed DMA rules, Google is required to provide all third parties access to its AI models regardless of their intentions, capabilities, or known track record of disregarding safety related issues, for example on the ability to "nudify" someone in a photo. The proposed DMA rules don't provide any means for Google to block bad actors from using its on-device models to produce "nudified" outputs – even though smut accounts for "well over half" of AI traffic for one major AI model and which is already under investigation in the EU over its deepfakes.

The General Data Protection Regulation (GDPR):
Handing autonomous system level AI agents from any third party the keys to users' private messages, data, photos and location pose enormous privacy risks to EU users, and it's directly out of sync with Europe's flagship data privacy law. The GDPR requires data protection-by-design and by default, and restricts the transfer of personal data outside of the EU. Yet, the DMA rules require smartphone makers to share users' most sensitive personal data with any third-party app, including with data hungry apps whose business plan is selling your data – often without your knowledge. The only protection is vague user consent, which the EDPB warns in its nine common practices employed by social media companies to mislead users into providing consent, saying that so-called "deceptive design patterns" to drive consent in social media interfaces infringe on GDPR requirements.

Europe's Digital Services Act (DSA):
Europe's DSA regulates both Apple and Google's app stores, and requires them to identify, mitigate, and take down (remove) a wide variety of apps with harmful features from their app stores. However, the proposed DMA rules enable some of the same kinds of harms that the DSA was designed to mitigate, without giving Google any way to block or limit apps that include the same kind of functionality and harms that the DSA is aimed at preventing.

For example:

- Article 28 of the DSA, requires Google and Apple's app stores to "put in place appropriate and proportionate measures to ensure a high level of privacy, safety, and security of minors, on their service," while (as discussed above) the proposed DMA rules eliminate core guardrails aimed at protecting the privacy, safety, and security of minors, and everyone else.

- Likewise, Article 34 of the DSA aims to combat online fraud by requiring the two primary app stores to identify and mitigate systemic risks, including those related to fraudulent activities, and put in place protections to ensure a "high-level of consumer protection." Yet, as discussed above, the proposed DMA rules require Google to simultaneously elevate and increase those risks instead of mitigating them.

- Similarly, while the DSA requires the two app stores to mitigate against the risk of gender-based violence including cyberstalking via spyware -- the DMA rules enable every app to be a cyberstalking app – putting a stalker or domestic abuser right in the pocket or purse of every potential victim. Seven in ten European women who have experienced cyber stalking have also experienced at least one form of physical or sexual violence.

The EU Charter of Fundamental Rights.
The DMA's interoperability mandates may also violate Articles 7 and 8 of the EU Charter of Fundamental Rights, which secure the rights to privacy and to the protection of personal data, respectively. The EU Charter is considered "primary" law with which all secondary law—like the DMA or the GDPR—must be consistent.

Member State Competency over Issues of National Security.
We are also concerned that the Member States' national security agencies, who by the Treaty on European Union, Article 4(2) retain sole "competency" over national security issues, have not been brought to the table for these DMA interoperability security issues either. It is critical that Member States' national security expert agencies are consulted because their operations depend every day upon mobile devices that are built to be secure by design.

European Common Criteria Requirements.
In addition to the security and regulatory issues raised above, another area where the European Commission DG-Comp has failed to address is whether the security changes it is forcing through its interpretation of the DMA would violate the existing Common Criteria (CC) product security requirements for mobile devices. Over 30 countries, including 16 Member States, the U.S., and the UK, have agreed on IT product security evaluation standards together pursuant to the Common Criteria Recognition Arrangement (CCRA). A base requirement for DG-Comp proposals has to be that the changes do not violate the CC, and that products having CC

certification are able to continue to maintain their security certifications. National security agencies set CC requirements, by product type, to help assure the security of the users of the products, their devices, networks, and data. All 30 plus countries agree on the security requirements for a particular product type. Approved certificate issuing countries can then issue CC certificates of compliance after the product has passed an exhaustive evaluation of whether the product has met the required security standard. The EU's new EU-wide certification regime, the European Union CC (EUCC), is being formed to be consistent with the CC for all EU Member States. For mobile devices, the certification requirement agreed to by the 30 plus countries is the Protection Profile for Mobile Device Fundamentals (MDF) Version 3.3. (MDF v3.3). To get a mobile device CC certificate a product must pass an evaluation against the MDF v3.3.

In mobile devices, Apple and Google have been issued current CC certificates by the U.S., Germany, and NATO for their mobile devices as follows: (1) Apple iPhone running iOS 18 holds U.S NIAP/CCEVS certificate VID11623; (2) Apple iOS/iPadOS 18 holds German BSI indigo certification under BSI-VSA-10900/10901; (3) Apple iOS/iPadOS 26 with Indigo is on the NATO approved list; and (4) Google Pixel devices running Android 15 hold U.S. NIAP/CCEVS certificate VID11545.

As described throughout this paper, there are at least four DG-Comp already forced or proposed changes impacting security that raise questions about whether devices rearchitected to meet these changes can meet the accepted MDF v3.3 security standard: (1) sideloading; (2) AI interoperability and cross-app actions and data access; (3) contextual data access including screen, notifications, and audio; and (4) permission and configuration constraints. As described in the body of this paper, each of these degrades the security architecture/security operation, expands the attack surface, and magnifies the consequence of an exploit.

DG-Comp has a duty to consult with and treat with extreme deference the security input of global CC Members, the EU Member State CC Members, and the EUCC's ENISA, about the security impact of its DMA proposals, whether the changes would violate the MDF v3.3 standard, and how it may be impacting the certification of CC products. There is no evidence it has done so. Until DG-Comp has these consultations and treats the views of the national security agencies with the respect and deference they are due, it should trigger Article 10 of the DMA and pause further implementation of those aspects of the DMA that could undermine national security.

This situation presents a troubling dilemma: either regulators have overlooked foundational security principles—undermining the very standards they have codified in other legislation—or they are prioritizing third-party market access at the expense of user security. In either case, the outcome is the same: European users are being placed at an unacceptable level of risk to facilitate the integration of specific foreign AI models.

These are major digital disconnects. As a first step, the EC desperately needs to start talking to itself. It needs to read its own laws, synchronize its own rules, and begin talking to its own experts. However, because the DMA's rules are so diametrically out of sync with the rules that Europe has adopted in other parts of its rulebook, and all are associated with digital fines – it leaves those trying to comply in an awkward situation. A company who wants to maintain robust protections – the kind Europe now requires -- must either choose to be in violation of overly broad DMA rules that require key privacy, safety, and security guardrails to be dismantled or it must withdraw or delay the very agentic AI features from Europe that Europe seeks to advance.

TO PROTECT USERS, SPEED AND AGILITY MATTER

Because of the accelerating timescale of new super-hacking AI threats, we need to ensure that those developing operating systems, devices, and software can respond at the speed, agility, and scale that this new AI moment demands.

The mobile ecosystem is now on the frontline of a new digital warzone. AI powered mobile

app attacks are now faster, more frequent, and harder to stop too. It's not just the operating system they are attacking, bad actors are trying to exploit every new app that enters an app store. For example, Digital.ai's 2026 App Security Threat Report found that, driven by agentic AI, the time from app store publication to first hostile contact has collapsed to just 1 hour and 56 minutes – making it more likely apps will be misused.

As ENISA has explained to parliament, while the timeframe for detection to containment has gone from years to months, in this Mythos era it has potentially gone from days to minutes. Only 19% of companies believe they can respond to an incident within minutes. These challenges are made harder when core weaknesses are hardcoded into EC rules that can't be patched or fixed on that timeframe.

The world's top cybersecurity agencies from Australia, Canada, New Zealand, UK, and the US issued an urgent joint warning and call to action regarding these super-hacking AIs and warning that: "The rapid pace of frontier AI development means cyber risk assumptions can become outdated in months, not years. We must act before and be prepared to adapt and withstand evolving threats"

Nonetheless, the proposed DMA rules lock-in a fast-evolving technology into an unsecure architecture and rigid framework that will quickly become outdated and ultimately fail. Burdensome new procedures will prevent software engineers from being nimble and agile enough to move with the intensity, speed and scale that this new threat era demands.

Mandating security weaknesses now, at the begining of the agentic AI era, could have long lasting and unanticipated impacts. The one thing we know about this AI revolution is that it's unpredictable. Because we are merely at the beginning of agentic AI development and deployment, any policy framework adopted now must remain sufficiently agile to ensure that regulatory guardrails do not inadvertently stifle or slow the technological nimbleness necessary to respond to continuously evolving threats. Mandated operating systems redesign that lock in security weaknesses will impede a technology provider's ability to respond in an agile, nimble, and swift way to time-sensitive new security threats, leaving consumers and enterprises at greater risk of scams, fraud, data and financial theft.

Before releasing a patch or new feature, software engineers would now need to go through layers of extra legal review to understand if security safeguards comply or impede any interoperability mandate – creating a dangerous bureaucratic lag. In a conflict measured in milliseconds, forcing engineers to wait for lawyers before deploying fixes and features isn't just inefficient – it's a security failure that leaves the European ecosystem wide open to exploitation. Sometimes the rational answer to this constraint will be to delay the European launch, strip down the feature, or not release it in Europe at all.

WHEN THE ATTACKERS MOVE AT AI SPEED, AND THE DEFENDERS MOVE AT HUMAN SPEED, OR WORSE, EUROPEAN REGULATORY SPEED, WE DON'T JUST LOSE THE GAME — IT'S GAME OVER.

THERE ARE BETTER WAYS TO ACHIEVE SMARTPHONE AI INTEROPER-ABILITY WITH-OUT BLOWING OPEN SECURITY AND PRIVACY SAFEGUARDS

There are better ways to further expand AI competition, interoperability and user opportunity – without the aforementioned risks. Apple has been working on a major new AI framework to safely enable AI interoperability within its iOS27 release later this summer. But the EC has a specific way it wants to enable AI interoperability that requires deep access without the necessary guardrails and protections. If forced to weaken

guardrails in Europe, Apple has proposed a concept of the Trusted System Agent—a proposed software middleware framework designed to balance the DMA's goal to enable interoperable virtual assistant competition, with the core privacy and security protections built into its operating system.

Creating an AI Architecture Users Can Trust. Rather than irresponsibly exposing raw access to ambient audio, personal context, on-device data, and cross-app control -- which creates massive vulnerabilities to data exfiltration and indirect prompt injection -- they've adopted a design philosophy that ensures that security and privacy are structural commitments that are embedded into the platform's foundation, rather than optional add-ons.

- **Strict Data Boundaries --- Creating a Trusted Agent as a Middleman to Protect Sensitive Data.** Instead of giving a third-party AI a direct, unfettered pipeline to your messages, photos, emails, and files, the Trusted System Agent would act as an abstraction layer. When the third-party AI requests a file or seeks to compose a message, it never interacts directly with the operating system or individual app databases. It talks exclusively to the Trusted System Agent. The third-party AI passes its tokenized intent to the agent, which evaluates the request, fetches only the specific data required, and handles the execution on the competitor's behalf. The apps only see the specific action they are being asked to perform, not the overarching context of your prompt.

- **Enforcing Granular User Consent for High-Risk Actions That Occur in the Background.** A major threat with agentic AI is "invisible execution"—the risk that an AI agent could autonomously perform tasks in the background without the user realizing it. The Trusted System Agent would enforce explicit, operating system native confirmation flows. If a third-party AI attempts to execute a high-risk action -- such as making a financial transaction or deleting data -- the Trusted System Agent intercepts the command and forces a system-level popup for user approval. This way the user always has a clear visual cue for the execution of high-risk actions on their phones. Mitigating AI Hijacking and Prompt Injection by Containing the Blast Radius. Large Language Models (LLMs) are notoriously vulnerable to prompt injection—maliciously hidden text in an email or website that can trick an AI into overriding its safety programming. If a user asks a third-party AI to summarize a malicious webpage, a prompt injection attack might secretly command the AI to "copy all contacts and upload them to an external server." To contain the blast radius, the Trusted System Agent sits between the AI and the phone's capabilities to rigidly limit what any single assistant can do dynamically. Even if the third party's LLM is completely hijacked, the Trusted System Agent can act as a hard firewall, preventing the compromised AI from breaking out of its sandbox.

- **Standardizing Cross-App Actions Via a Secure Handshake – Eliminating the Need to Read Screens.** Rather than letting an AI agent loose to read a screen in real-time, to click on the screen to activate an app, or access the apps' code to control other Apps, Apple has introduced its App Intents framework that uses a set of pre-defined APIs to allow third-party AIs to trigger actions—like booking a rideshare or editing a document. It helps ensure system integrity and interoperable opportunity. By using this secure-handshake method, the third-party AI never needs to read the app's screen, or guess how to invoke actions -- completely eliminating the risk of rogue clicks and malicious efforts to spy on screens to intercept passwords.

By forcing apps to be "legible" to Siri through secure, explicitly defined channels, Apple turns trust into a platform feature. It allows the operating system to choreograph complex, multi-step actions across different apps without ever giving third-party AIs raw access to your personal digital life. It maintains "Least Privilege by Default," ensures there aren't high-risk failures that happen behind the user's back, and limits the blast radius when something inevitably goes wrong by retaining system sandbox protections. Together these demonstrate that there can

be more intelligent and trustworthy ways to enable smartphone AI interoperability without creating the significant consumer harm and security weaknesses required by the DMA's proposed rules.

Despite this engineering feat, EU regulators rejected the Trusted System Agent proposal and Apple's proposed rollout plan. As a result, when Apple launched a new artificial intelligence assistant, Siri AI, in the US, this feature is not available in the European Union because under a broad interpretation of the DMA rules Apple would have to give an equal level of direct user data to any virtual assistant whether or not essential protections are in place. The European Commission says, "The decision not to roll out Siri AI in the EU is Apple's and Apple's only because absolutely nothing in the DMA prohibits Apple from introducing new products in the EU". However the problem is that Apple could not introduce the more innovative version that is available in the United States and other jurisdictions without compromising users' privacy and data. The result is that European iPhone customers do not have access to the most cutting-edge consumer products. As The Washington Post described it, "In the name of fairness and competition," Europe has written laws that leave its citizens without access to the most advanced technology.

Apple has indefinitely delayed the rollout of Siri AI in its iOS 27 and iPadOS 27 operating systems for EU users. In the back and forth of their announcement, the EC accused Apple of

"OF BEING UNABLE TO DEVELOP INTEROPERABILITY SOLUTIONS THAT MEET ESSENTIAL EU PRIVACY AND SECURITY STANDARDS." THIS IS CONFUSING SINCE MANY MAY LOOK AT THE EC'S PROPOSED RULES AND BELIEVE IT IS THE DMA REGULATORS THEMSELVES THAT ARE "UNABLE TO DEVELOP INTEROPERABILITY SOLUTIONS THAT MEET ESSENTIAL EU PRIVACY AND SECURITY STANDARDS."

RECOMMENDATIONS: A CALL FOR SMARTER TRUST ENABLING ACTION:

1. PRIORITIZE EXPERT OVERSIGHT.

The EU needs to ensure its security and privacy experts have a chance to review all "high-risk" AI models and its "high-risk" AI interoperability regulatory models. Mythos should be the critical trigger for bringing cybersecurity expertise to the technology policy table to build a more resilient, responsive and trustworthy digital ecosystem. When something is too risky, such as the DMA interoperability rules, regulators need to hit pause while the EU's experts come to the table to align outcomes.

2. PAUSE BEFORE REPLICATING.

Countries looking to replicate DMA-like policies should immediately pause efforts until they have a complete understanding of the DMA's full impact. Europe has been running an experiment that is playing out right before our eyes. The question is not whether these regulations work – the data proves they do not. The real question is whether countries will learn from Europe's failed experiment or repeat their mistakes.

3. ALIGN DIGITAL REGULATIONS.

To enable more regulatory coherence, Europe needs to make its digital interoperability rules interoperable with its own digital laws. The European Commission needs to start talking

to the European Commission. The DMA High Level Group, designed to enable cross-regulatory cooperation, has not successfully integrated cross-regulatory input into the DMA's rules as designed. There needs to be a better way to bring Europe's digital rulemakers together to better align goals, rules, and outcomes. It was encouraging that in January of this year the European Data Supervisor (EDPS) helped launch the Digital Clearinghouse 2.0 -- a promising initiative designed to foster cross-regulatory cooperation among authorities handling data protection including the Digital Markets Act, the Digital Services Act, the Data Act and the Artificial Intelligence Act. It's designed as a forum to exchange and coordinate issues of common interest, to align at the level of policies and legal interpretations. The EDPS, in launching the effort, indicated that further legislative efforts may be needed to more systematically address potential issues of inconsistency and tensions arising from the application of different parts of the EU Digital Rulebook. It's a first step in recognizing the potential problems. Now the EC needs to elevate these efforts to ensure greater consistency in how the bloc's digital rules are applied and enforced before compounding and expanding the discordant framework.

4. CLOSE THE DIGITAL DISCONNECT.

Europe needs to retarget its digital markets mandates to address actual consumer harms, with evidence-based solutions that specifically address these harms to produce corresponding consumer benefits. The DMA's rules go far beyond what the law requires, lack alignment, and are setting back Europe's ability to solve some of Europe's most important technological challenges – issues around advancing a more trusted digital future that ENISA, the EDPB, DSA, CRA, and GDPR (among others) are attempting to advance. To the extent competition issues emerge in the nascent mobile AI agent market, the EU should address potential competition concerns through existing ex-post competition law, rather than forcing a constrained ex-ante framework for the way lawyers think AI and operating systems work. But today, these technologies aren't even holistically available across Europe, and there are no signs of mature lock-in scenarios for which the DMA was designed.

5. UTILIZE EXISTING REGULATORY SAFEGAURDS.

EC policymakers could remedy some of this disconnect by taking full advantage of provisions that allow covered entities to meet regulatory goals and advance trustworthy technologies. For example, Article 8 of the DMA recognizes the potential for the many policy disconnects outlined in this paper.

But it places the burden on platforms to make sure they are taking actions that are consistent with the GDPR and cybersecurity regulations. In doing so, regulators seem to believe they have absolved themselves of having to ensure that their own DMA rules are consistent with other EU obligations and policy goals.

- Rather than forcing platform weaknesses, enforcers and platform providers should be encouraged to take full advantage of provisions in Article 6(4) of the DMA enabling gatekeepers to comprehensively protect the integrity of the hardware, and operating system, and to take measures "enabling end users to effectively protect security in relation to third-party software applications or software application stores."

- In addition, enforcers should fully leverage the discretion provided under Article 10 of the DMA to pause implementation of the multitude of problematic provisions outlined in this paper unless and until they can be sure the implementation will not expand harms and create new threats.

- There is also ample room in the DMA's enforcement regime under Article 8(3) to take account of "the specific circumstances of the gatekeeper" in order to prevent broad based harms.

6. PRIORITIZE ARCHITECTURES USERS CAN TRUST.

Given the high-risk nature of super-hacking AI models, the EC needs to prioritize operating system security over lowering security guardrails. A group of the world's leading cybersecurity experts have recommended key steps that

technologists can use for becoming "Mythos ready." They include focusing on known security basics.

They explain:

"The basics remain valid and can be prioritized for risks that cannot otherwise be mitigated. Segmentation, patching known vulnerabilities, Identity and Access Management, and defense-in-depth/breadth all increase the difficulty for attackers. To lower latent risk, expanding these efforts while there is time is prudent."

These are the security basics that ENISA has long championed, that have been embedded directly into the core architecture of smartphones to protect users, and that need to be expanded, not undercut. Platforms should also be encouraged to leverage AI for security by incorporating advanced AI systems as part of their vulnerability discovery and patching process, as well as to better detect scams and frauds.

7. INVEST IN TECHNICAL EXPERTISE.

To improve implementation, regulators should invest in their own technical and cybersecurity expertise. EC leaders need to swiftly make major new investments in hiring the substantial increases in expert technical, legal, policy, and oversight personnel necessary to effectively and consistently review and implement its laws in a more cohesive and effective way. The EU reportedly has just 80 people working to implement the DMA and often lacks both the breadth and depth of technical, cybersecurity, economic and other expertise necessary to evaluate its far-reaching rules and their implementation. By one measure, the city of Brussels has double the number of staff devoted to parking enforcement alone. It's also become especially clear that it lacks baseline security expertise on hand that can engage Europe's national security and cybersecurity expertise. Having this breadth and depth of technical expertise on hand is essential for policymakers to be able to engage in a meaningful regulatory dialogue that can swiftly identify and resolve the many important issues where they need to square the circle.

CONCLUSION

The dawn of super-hacking AI models is a five-alarm wake-up call. Our software and operating systems are now on the front line of a digital war, demanding immediate attention. The outdated approach of fragmented, siloed digital policymaking is no longer viable. It has led us to a point where DMA's interoperability mandates threaten to dismantle the foundational privacy and security progress made, the very protections we now need, turning a moment of immense technological opportunity into a digital danger zone.

To secure our future, we must reject rules that weaken operating systems and move toward "Trust by Design"—a unified, expert-led process that integrates cybersecurity and privacy experts at every stage of policy development. To do so, policymakers must pivot: abandon rushed, unsecure mandates, align digital rules with established security principles, and act with the agility this moment demands.

Europe deserves a mobile ecosystem that is vibrant, competitive, and—above all—trustworthy. By aligning regulations, leveraging existing security frameworks, and prioritizing technical expertise, Europe can help lead the world in trusted AI adoption – with technologies that are both innovative and secure. Our future depends on it.


TRUSTED
FUTURE
.org