Episode 239 – How Is Space Law Adapting to Autonomous Missions?

Speaker: Jesús Bernal Allende, Space Lawyer and Founder of IURUS Consulting – 22 minutes

John Gilroy:

Welcome to Constellations, a podcast from Kratos. My name is John Gilroy and I'll be your moderator. Today, we're looking at how existing space law is being applied as autonomous systems take on a larger share of spacecraft decision making from actions not directly commanded to situations where operators can't always spell out how a system reached a choice. Joining us is Jesus Bernal Allende, space lawyer and founder and director of IURUS Consulting to discuss how responsibility, licensing and coordination are being interpreted as autonomy becomes part of modern mission design. Welcome, Jesus.

Jesus Bernal:

Hi, John. Thanks for having me. I'm glad to do this.

John Gilroy:

Jesus, from a legal standpoint, where are you seeing the biggest gaps as autonomy becomes more common in space operations?

Jesus Bernal:

This is the one I get asked the most and for good reason. There are four gaps I keep coming back to and none of them are really about the technology itself. The first is the agency gap. Right now the law only knows two categories, a tool or a legal person. An autonomous system that makes its own operational calls doesn't fit either box. In my own work, I call this the tertium genus problem, the third category nobody's built yet. Second is the supervision gap. Article six of the Outer Space Treaty says, "States need continuing supervision of their space activities." That made sense when a human was checking in regularly. It doesn't hold up when you've got thousands of satellites making their own calls faster than any human could review them.

Jesus Bernal:

Third, the responsibility gap. The liability convention runs through states and that framework still holds. But once you have private operators and autonomous subsystems all contributing to the same incident, the chain of causation, the law was built to trace just breaks down. And fourth, the systemic damage gap. The whole legal architecture was built for one incident, one victim. It was never built for accumulating orbital debris or a cascade of conjunction events across an entire constellation. It's not that the law got it wrong, it's that it was built for a world that doesn't really exist anymore. And the fix isn't more rules, it's new categories.

John Gilroy:

When an autonomous system takes an action that wasn't directly commanded, what does current law actually allow us to do with that fact?

Jesus Bernal:

Here's the part that surprises people, it doesn't matter. Legally, it makes zero difference whether a human pushed the button or an algorithm made the call. Under the Outer Space Treaty and the Liability Convention, responsibility attaches to the activity and to the space object, not to whatever decided to move it. For strict liability, the only question is whether damage happened. Saying the algorithm made the call doesn't change that. For fault-based liability, once there is no human command to point to, the analysis shifts, design choices, deployment decisions. Did the operator supervise at the level the system actually required? The human is still in the picture just earlier in the timeline and that's where it gets more interesting.

Jesus Bernal:

The way I evaluate this is through what I call functional intentionality. You look at what the system was optimized to do, not its intent in any moral sense. Could the operator have anticipated this outcome knowing how the system was designed? That's the question that actually holds up in court. And functional intentionality gives you a way to answer it without having to prove what the system intended, which you never could. You look at the design choices, the optimization targets, the deployment context. If the outcome was consistent with what the system was built to do, the operator had the means to foresee it. That's enough.

John Gilroy:

How should responsibility be handled when an autonomous spacecraft's behavior affects another operator's mission intentionally or not?

Jesus Bernal:

The instinct is to find one responsible party and assign the blame. That instinct breaks down fast when you have multiple autonomous systems, multiple operators and multiple jurisdictions all contributing to the same incident. What you need instead is layers. And I built what I call a five layer accountability architecture for exactly that. The first layer is the system itself. There's a pre-agreed guarantee fund that pays out automatically for demonstrated harm. No need to wait for blame to get assigned. No need to prove fault before someone gets compensated. The fund moves first. The legal process catches up. The second layer is the operator, supervision, maintenance, making sure the system is integrated properly into the broader safety case.

Jesus Bernal:

The third is the designer where the analysis shifts to design defects or capabilities that don't match what was advertised. It keeps going up the chain from there. The fourth layer is the certifier. If the certification approves something it shouldn't have, that approval carries consequences. And the fifth is the regulatory framework itself. If the standards were clearly inadequate given what was known at the time, the people who wrote them are not exempt from that conversation. The point of the architecture is proportionality, not punishment. Each layer absorbed the part of the problem it actually created. And on enforcement, the more durable approach is infrastructure over sanctions, certification requirements, audit rights, transparency obligations. Sanctions arrive after the harm is done. Infrastructure changes what's possible before that point and it travels across jurisdictions a lot more reliably than threats do.

John Gilroy:

How are space lawyers thinking about responsibility when operators can't fully explain how or why an onboard system made a decision?

Jesus Bernal:

The key move is separating two things people usually treat as one. The first is explainability, putting the decision into plain language. The second is traceability, the ability to reconstruct exactly what happened. The logs, the model version, the inputs, the test data. Explainability is not always possible with complex models, especially under time pressure. Traceability has to be non-negotiable. Even when the system can explain itself in the moment, it still has to leave a verifiable paper trail. And this is where it gets practical. I work with a framework I call evidence sovereignty, five conditions that give data real legal weight.

Jesus Bernal:

Each one answers a question a court would actually ask, "Where did this data come from and can you prove it? Will the same inputs produce the same result if you run it again? Can the decision be tested against a competing hypothesis? Was the evidence reviewed by someone with no stake in the outcome? And was the data still valid at the moment it was used or had it already gone stale?" Those five together are what make a log legally defensible, not just technically complete. The part operators should actually care about is this. If the logs are missing or corrupted, every decision made during that blackout gets presumed invalid until proven otherwise. Sloppy logging stops being a paperwork problem and it starts being a liability problem. Explainability is a design aspiration. Traceability is a legal obligation. The two are not the same and confusing them is where most operators get into trouble.

John Gilroy:

On that wavelength, how are lawyers approaching risk assessment when machine behavior can vary in ways that aren't fully predictable?

Jesus Bernal:

The smarter move is containing behavior instead of trying to predict every outcome. Prediction at that level of complexity isn't as in though you never reach. Containment is an engineering constraint you can actually build and verify. And three mechanisms are doing most of that work right now. The first is behavior containment. The system is designed so that even surprising behavior stays inside a safe envelope. The value is that you don't need to anticipate every scenario. You just need to know the system can't leave the envelope no matter what it encounters. That shifts the legal question from, "Did we foresee this?" To, "Did we build the boundary correctly?" The second works as an early warning layer. Certain conditions automatically trigger a stricter regime. More logging, more checks.

Jesus Bernal:

The system flags itself for closer scrutiny before a human even knows something unusual is happening. That's what activation thresholds are built to do. And in high latency environments where a human response takes minutes, that automatic escalation is the only supervision that actually arrives in time. And the third is an independent subsystem that watches continuously for those threshold violations. It's

called a dignity guardian. And this design principle is deliberate veto authority, not redirect authority. It can stop a decision. It cannot substitute its own. That distinction keeps the accountability chain intact and the authority exactly where it was designed to sit.

John Gilroy:

As spacecraft began acting with greater independence, especially far from earth, how do you expect today's licensing regimes to stretch or evolve to keep pace with that shift?

Jesus Bernal:

Licensing is quietly shifting from mission-based to architecture-based authorization and the regulators are ahead of most operators on this. The old model was, "Describe your mission, get a license, operate accordingly." The new model asks a different question. Does your system have the architecture to handle what is actually going to encounter? Can it make safe decisions autonomously when it has to? That shift changes what you have to demonstrate to get authorized in the first place. And this isn't theoretical. The FCC's 2024 orbital debris order already defines a satellite's end of life partly by whether it can perform autonomous collision avoidance.

Jesus Bernal:

And it points to the capability newly implemented in the Sterling constellation as the operational reference. On the launch site, the FAA pulled back its proposed 25-year disposal rule for spend rocket stages earlier this year. A separate process on a separate category of object. While the FCCC's tougher five year mandate for satellites, the one built around that exact capability based logic stands untouched. Two agencies, two objects, same underlying shift. Regulators aren't waiting for a treaty revision. They're already authorizing by architecture instead of by mission, one rule making at a time. The label is mine, but the practice is already theirs.

John Gilroy:

For missions where communication delays eliminate human-in-the-loop control, what forms of oversight are lawyers advising operators to establish on the ground and in system design?

Jesus Bernal:

Oversight has to stop being about the moment when there's no human-in-the-loop, the supervision has to happen before the decision and in the record it leaves behind. Those are the only two windows you actually have. Before means constraints and safety logic built into the architecture itself. So the default behavior is already safe regardless of what the system encounters. The record means immutable logs, sensor data, internal states stored in a way that can survive a cross-border audit years after the fact. And what you do with that record afterward, counterfactual analysis, third party review is where accountability actually gets built. And this is where the framework does something different.

Jesus Bernal:

For the highest stakes decisions, I designed something called the parliament of models. Multiple model configurations deliberating or voting inside a fixed governance structure before a critical decision executes. The redundancy is valuable, but what matters most is the trail it generates because that trail is

the only thing a regulator or a court will have to work with when no human was there to weigh in at the time. When the signal takes 20 minutes to arrive and 20 minutes to come back, you can't supervise the decision in real time. That's a physical constraint, not a governance failure. The governance question is whether the architecture was built to supervise itself and whether it left enough behind to reconstruct exactly what happened and why. Those are the two things operators actually control.

John Gilroy:

What challenges could arise as some operators deploy advanced autonomy while others continue to rely on older, slower operational models?

Jesus Bernal:

The problem runs deeper than any one operator. It's structural and it shows up in three places. The first is temporal. Some systems react faster than others, which means that in a shared orbit, one operator's collision avoidance window might close before another operator's system has even registered the risk. The speed gap is not just an operational inconvenience. In a conjunction scenario, it determines who had the ability to act and who didn't, which is exactly the kind of question that ends up driving liability analysis. The second is epistemic. Operators running more advanced autonomy accumulate richer data over time. That data informs better models which generate better data and the gap between fast and slow operators widens not because anyone is falling behind deliberately, but because the architecture compounds.

Jesus Bernal:

At some point, the two systems are operating on fundamentally different pictures of the same environment. And the third is about responsibility. When something goes wrong between two systems operating under completely different governance frameworks, attributing fault becomes genuinely hard. The architecture itself makes it difficult to know where one system's decision ended and another's began. That's not a procedural problem you can solve with better lawyers. It requires a different kind of infrastructure. Their fix is normative interoperability, shared traceability standards, common communication protocols so fast and slow systems can still talk to each other procedurally even when they're built completely differently. You don't need everyone operating the same way. You need everyone leaving a record the other side can read.

John Gilroy:

Where do you see the biggest disconnect between how government rules for space operations are currently written and the kinds of autonomous capabilities operators are beginning to field?

Jesus Bernal:

It comes down to four assumptions baked into the rule book. And the problem isn't that they were wrong when they were written. The problem is that the world they described no longer exists. The first continuous human supervision as the default. That made sense when missions were few, slow and close enough to earth for real time communication. It doesn't hold up for deep space operations or high tempo constellations where decisions happen faster than any human review cycle. The second is the assumption that the operator and the responsible party are always the same entity. In practice, design,

operation and infrastructure control are split across companies, contractors and jurisdictions. The chain of responsibility, the law was built to trace runs through a structure that rarely exists as described. The third is human timescales. The law was written assuming a human could review a decision before it mattered.

Jesus Bernal:

That assumption breaks completely when you're dealing with millisecond decision making in a conjunction scenario where the window to act is gone before the alert reaches the ground. And the fourth is the one that touches everything else. The entire framework assumes that only states act on the international stage. Today, private operators and autonomous systems are doing most of the operational work and the rule book has no category for either. A private constellation making 10,000 autonomous decisions per day is not a state activity in any meaningful sense and treating it as one produces the wrong answers at every level. What you need instead is a functional taxonomy of roles, validation tied to actual performance and legitimacy grounded in evidence rather than authority. The common thread is moving away from who you are on paper and toward what you can actually demonstrate. That's the only logic that holds up when the actor making the decision isn't a state, isn't a person, and isn't operating on any timeline a human could supervise.

John Gilroy:

From a regulatory perspective, what demonstrations or performance data will be needed before autonomous spacecraft are treated as reliable actors in their own right?

Jesus Bernal:

I'd actually push back on the framing a little. The goal isn't really to call the system itself reliable. Reliability is a property of hardware. What we actually need to evaluate is whether each individual decision the system produces holds up to scrutiny. That's a different question and it leads to a different kind of evidence. The evidence breaks down into four categories. The first is architectural. Does the design match a grid safety norms? The second is traceability. Can the decision be reconstructed and audited after the fact? The third is performance under critical conditions, not just how it behaves on a good day when nothing is at stake. And the fourth is institutional learning. Are incidents actually getting fed back into the system and the standards? Or does each failure get treated as an isolated event?

Jesus Bernal:

And here's where the framework does something different. Not every system needs to hit the same bar and that's by design. I use something I call the regulatory cube. It calibrates how much evidence you need based on two intersecting variables, how critical the mission is if something goes wrong and how much of the decision making the system handles without human input. A small satellite in a stable orbit with limited autonomy gets evaluated differently than a deep space system making irreversible decisions with no human in the loop. Higher stakes, higher bar, lower stakes, lower bar. But every system gets held to some appropriate standard and the cube tells you which one.

John Gilroy:

Looking ahead, what legal tools, standards or shared norms do you think operators will need to support autonomy at scale across commercial, civil and government missions?

Jesus Bernal:

The real work here is institutional and it comes down to five pieces that operators will need before autonomy at scale becomes governable. Universal registration and noted going beyond just tracking objects to actually tracking technical profiles and governance metadata so you know not just where something is, but how it's governed and what it's authorized to do on its own. Systems change after their approved updates, retraining, new deployment context. Continuous validation means reassessing periodically instead of signing off once and walking away. The approval process has to keep pace with the system, not just clear it at the start. When something goes wrong, the question of who carries the weight has to track the decisions that were actually made. Proportional responsibility means matching liability to the level of autonomy and risk the operator chose to deploy. More decision making authority, more legal weight when that authority produces harm.

Jesus Bernal:

Conflicts between autonomous systems operating under different frameworks are going to happen. Deliberation mechanisms give you a structured way to resolve them before the conflict becomes an incident. Right now, mostly that procedure doesn't exist and safeguards against power concentrating in too few hands, whether that's data, infrastructure or the standards themselves. The entity that controls the logging standard controls what counts as evidence. That's not a technical detail. That's a governance question. Most of this is infrastructure nobody thinks about until something breaks. And when it breaks at a scale in a space, the window to fix it is a lot narrower than it is on the ground. Without traceability, there's no accountability. Without accountability, there's no trust. You want that plumbing in place before the incident that makes it unavoidable, not after.

John Gilroy:

Which aspect of autonomy do you think is most likely to generate regulatory or operational friction as missions become more complex?

Jesus Bernal:

It's not any single factor. It's what happens when a few of them line up at once. The first is divergent standards. Operators following completely different safety norms and logging practices, which means there's no common language when something goes wrong. You can't reconstruct what happened if the records weren't built to talk to each other. The second is different iteration and speeds. Some teams updating their systems constantly. Others barely touching theirs, and the gap between them creating coordination blind spots, neither side fully sees. The fast operator assumes the slow one will keep up. The slow operator assumes the fast one is still behaving predictably. Neither assumption holds. And the third is emergent on predictability.

Jesus Bernal:

This one is different in kind, not just degree. It's not that one system behaves unexpectedly. It's that multiple autonomous systems, each behaving exactly as designed, interact in ways nobody individually

planned for, and produce outcomes that weren't in anyone's risk model. No one made a mistake. No one was negligent. The architecture itself generated the problem and that's precisely what makes it so hard to assign responsibility afterward. The first major incident with unclear attribution is what actually triggers the reckoning. That's when the gaps in jurisdiction and procedures stop being theoretical and start being front page news the same way it has played out in pretty much every other high tech domain before this one. What gets you through that without backing away from autonomy altogether is having the governance architecture build ahead of time. Dispute mechanisms, registries, safeguards, not drafted in a panic afterward.

John Gilroy:

Jesus, I think you've given our listeners a better understanding of the legal challenges of autonomous spacecraft. I'd like to thank our guest, Jesus Bernal Allende, founder and director IURUS Consulting.

Jesus Bernal:

Thanks, John. Glad we could get into it. The future of a space is autonomous either way. The only real question is whether the governance gets to keep up with it.