

MILESTONE EVALUATION / PRE-LAUNCH GATE

007 First Light

Pre-Launch Evaluation - Sample Report

About this sample. This is a demonstration piece, written in July 2026 after 007 First Light shipped, to show the format, depth, and kind of judgment a client receives. It is written as it would look at a pre-launch milestone gate, using only what was knowable before the game's May 27, 2026 launch, and it ends with a scorecard comparing its calls against what actually happened. Real engagements happen pre-release, on the client's build, under NDA. 007 First Light was chosen as a subject I have no professional connection to; this document is not affiliated with or endorsed by IO Interactive.

Basis of evaluation: completed full playthrough on PlayStation 5. **Evaluator background:** 18 years in game development, including eight years at Square Enix reviewing milestone builds across a AAA portfolio.

Executive Summary

007 First Light arrives with a built-in expectations problem: it is sold, on the box and in every trailer, as a James Bond game from the creators of *Hitman*. Judged as a cinematic action-adventure, it is excellent - the studio's best story work to date, a surprisingly strong combat system, and production values that stand up against anything in the genre. Judged as an heir to *Hitman*'s design philosophy, it disappoints, and that gap is where the risk sits.

The core issue is player agency. Outside its best infiltration sequences, the game often sets one approved path through a situation and actively resists everything else. The player is regularly a passenger in their own mission: permission to kill is switched on and off at the designers' leisure, improvised solutions are blocked or redirected, and a plan going wrong means a game over screen rather than the messy, memorable recovery that made the studio's name. None of this will stop the game reviewing well or selling. It will show up as the consistent caveat in mid-range reviews, cap replayability, and become the loudest complaint in player discussion once the launch honeymoon ends.

This evaluation expects a Metascore between 80 and 85 and a very strong commercial launch. Recommendations at this gate focus on what can realistically still change before launch: how the game is positioned in the final marketing beats, the worst of the invisible walls in the early missions, how failure is handled, and an AI tuning pass.

WORKING WELL

- Ranged-to-melee combat flow, made better by scarce ammo
- Story, cast, and dialogue quality
- Infiltration sequences with multiple ways in
- Production values across the board

NEEDS ATTENTION

- Missions that dictate the path and take choices away
- Fail states that punish improvisation instead of absorbing it
- A final act that runs out of steam
- Patchy enemy AI; gadgets that wear thin

Client Questions

In a live engagement this section answers the questions the client sends in ahead of the review. For this sample, these are the four questions I would expect this build to raise inside its own team.

"Combat has never been this studio's strong suit. Does it hold up?"

Yes, and it is the build's biggest positive surprise. The worry was fair: heavy combat was the weakness of the *Kane and Lynch* games, and *Hitman* was never asked to carry long firefights. *First Light* clears the bar. Weapon feel is good rather than class-leading - it will not out-shoot the dedicated shooter franchises, and it does not need to. What sets it apart is the flow between shooting and melee. Ammo is limited to what a weapon carries, with no magic reserve pool, so the player has to scrounge the arena mid-fight, and the system rewards it: empty a pistol, throw it at an enemy's face, close the distance, finish the fight against a wall, pick up their weapon, and keep moving. That loop is dynamic, intuitive, and genuinely fun, and it is a pillar the reviews and the marketing can safely lean on.

"How will 'from the creators of Hitman' land with players?"

Right now it is a liability dressed up as a selling point. The phrase promises a specific kind of game: open problems, tools the player can combine freely, and plans that go wrong in recoverable, memorable ways. This game deliberately does not do that, and the audience most likely to buy on day one - the studio's existing fans - is exactly the audience that will notice. If expectations are not reset before launch, the gap between the promise and the product gets discovered in public, in reviews and forums, instead of being managed in the messaging. Recommendation 1 deals with this directly.

"Where will critics push back?"

Three places, in order of how much they will matter. First, the steering: reviewers will describe the feeling of being told what to do - permitted actions switching without warning, one approved route where the game's own systems suggest three. Second, the final act: the mysteries resolve early and the push to the ending feels padded; several reviewers will say the game runs out of steam. Third, believability: enemy AI is patchy enough that guards sometimes ignore a takedown happening next to them, and the gadget set loses its novelty through repetition. None of these sink the headline score; all of them will show up in the text of the reviews that matter.

"What should we expect at review time?"

A Metascore between 80 and 85. The production values, the story, and the licence pull it up; the agency problem and the weaker final act pull it down. There is no direct competitor - this is the first major Bond game since 2012, so there is real pent-up demand waiting for it. User scores will land below critic scores, dragged by the *Hitman* core. Full reasoning in Score Guidance, page 5.

Market Context and Expectations

The licensed action game has escaped its bargain-bin history. *Indiana Jones and the Great Circle* reset the expectation in 2024: a AAA licensed title from a specialist studio is now assumed to be good, and gets scored against the genre's best rather than against other licensed games. That is the company *First Light* will be judged in.

Comparable title	Year	Metascore	Why it matters here
Uncharted 4: A Thief's End	2016	93	The genre ceiling; the cinematic action benchmark
Alan Wake 2	2023	89	Story-led production values at the top of the band
Hitman 3	2021	87	The studio's own bar, and the source of expectations
Indiana Jones and the Great Circle	2024	86	The licensed revival benchmark; closest comparison
Star Wars Jedi: Survivor	2023	85	Licensed AAA action at scale
Marvel's Guardians of the Galaxy	2021	80	Proof that strong reviews alone do not guarantee sales
RoboCop: Rogue City	2023	72	The serviceable licensed floor

Scores as at April 2026, lead platform. Sourced from Metacritic.

What the market expects from this specific game

Three audiences are buying three different games. Bond fans want the fantasy: the tailoring, the gadgets, the wit, the set pieces. The general action audience wants a polished 15-hour cinematic campaign in the *Uncharted* mould. The studio's existing fans want the freedom to make their own plans, because the marketing keeps reminding them who made it. The build delivers on the first two convincingly. The third is the unmanaged risk, and it sits with the most vocal part of the audience - the people who write the forum threads and Steam reviews that shape how a game gets talked about after launch.

Commercially, the game is in a great position: a dormant mega-licence, a first trailer that led a PlayStation State of Play, a theme song from a global artist, and nothing competing in the launch window. The commercial risk of the agency problem is low at launch. The cost shows up later - in word of mouth, replay numbers, and how much goodwill a sequel starts with.

Strengths

The combat is better than the studio's track record

Covered under Client Questions, and worth repeating as a strength in its own right: the flow between shooting and melee, the hard-hitting melee system, and improvisation forced by scarce ammo add up to the best moment-to-moment combat the studio has shipped. The ammo rule deserves specific praise: Bond carries only what the weapon holds, which turns every long fight into scanning for weapons, moving, and taking chances, instead of sitting in cover.

The story work is the best in the studio's catalogue

The origin structure - stumbling into an MI6 operation, recruitment, training, then a disastrous mission that shakes MI6 and the 00 programme itself - gives the story a real arc rather than a mission list. The playable training montage is the standout idea: the player controls fragments of the montage as Bond visibly gets better, which is character building through gameplay rather than cutscene. The supporting cast lands: fellow trainees Cressida and Monroe and instructor Greenway are distinct and likeable, and M, Moneypenny, and Q deliver briefing dialogue that feels natural and plays well off the cast. This is some of the best narrative work in recent AAA, and it should lead the messaging.

The infiltration layer works where it is allowed to

When the game opens up, the studio's DNA shows. Scoping out a location, finding clues that unlock different approaches, choosing between climbing the outside of a building or pickpocketing an access pass and walking through the front door: these sequences are the best of both worlds, Bond fantasy delivered through real choice. They prove the team can still do this. The problem, covered opposite, is how rarely the game trusts them to.

Production values stand up at the top of the genre

Presentation, performance capture, locations, and audio all sit comfortably alongside the titles in the comparison table. Nothing here needs fixing at this gate; this is the foundation the high scores are built on.

Risks

R1. The passenger problem: the game takes choices away from the player

The build's central risk. The player is frequently not allowed to solve situations with the tools the game has already handed them. Permission to kill is switched on and off by the game - no licence to kill right now, so the loaded gun in your hand might as well be a prop - and the switch flips at the designers' leisure, not for reasons the player can understand. Many missions run on a single approved path with everything else blocked or redirected. Over a full playthrough it starts to feel like playing with a design team standing over your shoulder, telling you where to go and stepping in every time you show initiative.

The clearest example is how the game handles failure. In the studio's previous work, a plan going wrong was where the game started: alarms, improvisation, escalation, and the memorable recoveries that drove replays. Here, the same moments end in a game over screen. The systems that could make failure interesting are already in this build - the combat is strong enough to absorb a blown infiltration - but the mission design refuses to let them. This is a deliberate design choice, not a bug, and it is too late to reverse at this gate. The recommendations below go after its worst moments and the messaging around it, not the design itself.

R2. The final act runs out of steam

The pacing holds while the mysteries do. Once they resolve and the push to the final confrontation begins, the seams show: encounters that feel like padding, and a closing stretch that reads as if the team realised the well-paced story did not add up to enough playtime. Reviewers finish a game right before they write about it, so the last act counts for more than its share of the runtime.

R3. Patchy AI undermines the stealth

Enemy awareness is uneven to the point of comedy in places - guards blind and deaf to a takedown happening right next to them. In a game that requires stealth for long stretches, that does not read as difficulty tuning; it reads as the world not being real, which quietly cheapens every stealth win the player earns.

R4. The gadgets wear thin

The gadget set entertains on introduction and then repeats. You can only make a guard sick with a dart, slip past, or pickpocket someone so many times before it stops being interesting, and the game reaches that point by mid-game. The problem is the pacing of new toys and what you are allowed to combine, not the toys themselves.

Score Guidance

Measure	Expectation	Reasoning
Metascore	80 to 85	Production quality, story, and licence goodwill carry it into the band; the agency problem and the weak final act stop it reaching the genre's best
User sentiment	Positive, below critics	The studio's core fans show up early and loud in user reviews, and they are the group most likely to reject the design
Launch sales	Very strong; studio records plausible	Dormant mega-licence, nothing competing in the window, premium marketing beats already banked
Long-tail risk	Moderate	Low replayability plus a loud linearity debate will shape discount-window sentiment and how much goodwill a sequel starts with

Recommendations at this gate

Ordered by impact against what can realistically still change this close to launch. Structural redesign is off the table; framing, tuning, and messaging are not.

- **1. Reset expectations in the remaining marketing beats.** Stop selling *Hitman* adjacency and sell what the build actually is: the best interactive Bond story yet made. Every trailer that leans on the sandbox heritage sets up a promise the game does not keep. This is the biggest-impact change available and it costs no development time.
- **2. Audit the invisible walls in the first six missions.** Find the ten worst moments where the game blocks an action its own systems support, and where possible either allow the action or give the player an in-world reason it does not work. The early missions carry first impressions, reviews, and refund windows.
- **3. Soften the most common fail states.** Where a blown stealth sequence currently means game over, turn the most-hit cases into a fight or an escape instead. The combat system is strong enough to absorb it, and every conversion turns a frustration into a story players tell.
- **4. Run an AI tuning pass** on the encounters where guards ignore what is happening next to them, starting with the long stealth stretches where believability matters most.
- **5. Space out the gadgets.** Hold at least one back until mid-game and vary what can be combined, so the toolset stays interesting as the missions get longer.
- **6. Post-launch: serve the systemic audience deliberately.** A challenge or remix mode built on existing missions - looser rules, scored improvisation - is the cheapest honest answer to the *Hitman* audience, and a better one than pretending the campaign is something it is not.

The Scorecard: What Actually Happened

Written eight weeks after launch. This section exists because this is a retrospective demonstration: it shows the kind of calls a pre-launch evaluation makes, checked against the public record. See the note on page 1.

Pre-launch call	What happened	Verdict
Metascore 80 to 85	87 on PlayStation 5 (88 on Xbox Series X, 84 on PC) - among the studio's highest-rated releases	Miss - it did better
Very strong launch; studio sales records plausible	1.5 million units in 24 hours, publicly reported as the studio's fastest-selling game	Hit
Agency and replayability as the recurring caveat in mid-range reviews	Eurogamer (4/5): 'less cerebral and replayable' than Hitman. PCGamesN (8/10): 'doesn't feel as free-form and creative' as Hitman	Hit
Linearity becomes the loudest player complaint after the honeymoon	Within weeks, community discussion ('Linear or not?') settled on fixed paths, converging routes, and guidance markers as the core criticism	Hit
Combat as an under-appreciated strength	Reception split: praise for the fistfights and flow, while some outlets found the forced cover-shooter sequences clumsy	Partial

The score call was under, and it is worth being honest about why. I had the agency problem and the weak final act costing more with critics than they did. The licence, the presentation, and the plain relief of a genuinely good Bond game carried it two points past my ceiling on the lead platform. The calls about where the criticism would focus all landed; what I overestimated was how much that criticism would cost the headline number. That is a useful lesson in itself: critics review the game in front of them, and players review the game they were promised.

The partial is worth a note too. Critics who disliked the combat met it when the game forced it on them mid-stealth; the system is great when chosen and grating when imposed. That is not a contradiction of the pre-launch read - it is the agency problem showing up inside the combat reception, and it is exactly the kind of knock-on effect a milestone review exists to catch before the reviews do.

Public record checked 20 July 2026: [Metacritic](#); [Eurogamer review excerpt](#); [PCGamesN review](#); [launch sales report quoting IO Interactive](#); and public community discussion on Metacritic and Steam.

What a Real Engagement Looks Like

A live mock review is done on your build, at your milestone gate - first playable, vertical slice, alpha, beta, or pre-launch - under NDA. You get a written report in this format, scoped to the questions you send in, followed by a live debrief with your team. The evaluation covers what a publisher's gate review would cover: an honest read on quality, market and score context, and recommendations matched to what you can actually still change at your stage.

Peter Fury | peterfury.com/reviews | natehawke.struggles@gmail.com | +44 7541 230 521