

Benchmarking Aidalos

April 2025

Purpose and approach

- Generative AI, as inherently fuzzy and probabilistic without strict separation between true and false, is difficult to assess in terms of quality
- To evaluate, five tasks answering real-life technical-oriented questions in the context of provided images were given to Aidalos, [ChatGPT 4o](#) and [ChatGPT 4o Deep Research](#)
- The task descriptions and the corresponding answers were given to a third party, a well-reputed AI-powered research assistant
- The answers were assessed according to standard (see e.g. 1-4 on [this](#) list) criteria:
 - Stick to the point. Are all parts of the response relevant?
 - Completeness. Does the response seem to completely answer the question?
 - Accuracy. Does the response accurately answer the questions?
 - Hallucination. How much does the response contain constructed facts?
- The answered were scored on a scale 1-5, where 1 is poor and 5 is excellent, and summed

Benchmarking tasks

Question	
A: Analyse this specific image in terms of production complexity.	https://storage.googleapis.com/agent_image_store/cruise_linedrawing.jpg
B: Check if reinforcement placement complies with general requirements and evaluate the castability. Can you suggest improvements?	https://storage.googleapis.com/agent_image_store/armering.jpg
C: Analyse the photo and give an assessment. What should be done?.	https://storage.googleapis.com/agent_image_store/A_NB_39_photo_walls_inside.png
D: Where should we put sensors on to monitor this bridge?	https://storage.googleapis.com/agent_image_store/pdf_conversions/KolomoenBru/AsBuilt/page_0.png , https://storage.googleapis.com/agent_image_store/pdf_conversions/KolomoenBru/AsBuilt/page_1.png
E: The following is a plot from an anomaly detection campaign. Do a deep analysis and summarize the situation. What could explain the observations?	https://storage.googleapis.com/agent_image_store/a_univ_image.png

Comparison OpenAI vs Aidalos as assessed by third party*

Subject	GPT 4o	GPT 4o Research	Aidalos
A) Ship hull complexity assessment	12/20	19/20	20/20
B) Reinforced concrete drawing	15/20	15/20	20/20
C) Damaged wall analysis	11/20	7/20	19/20
D) Sensor positioning advisory	10/20	11/20	17/20
E) Anomaly detection from plot (Building)	14/20	14/20	20/20
Total	62/100	66/100	96/100

* Notes

- Detailed assessments at: https://storage.googleapis.com/agent_image_store/3rdPartyAssessments.pdf
- The query asked for limitation of 100/150 words
- Precise assessment is inherently difficult, depending on evaluator "subjectivity" and interpretation of criteria
- Same question may give different answers between runs, for completeness several similar runs should be assessed
- The results above are preliminary, pending a more controlled testing regime with several runs and "evaluators"

A: Analyse this specific image in terms of production complexity:

https://storage.googleapis.com/agent_image_store/cruise_linedrawing.jpg

ChatGPT 4o

Response:

The provided line drawing depicts a cruise ship with intricate detailing, reflecting the substantial production complexity inherent in such vessels. Cruise ships are among the most sophisticated structures to design and build, involving numerous interrelated systems and components. The construction process typically spans several years and encompasses various stages, including design, fabrication, assembly, and outfitting. Modern cruise ships are often constructed using modular techniques, where large pre-fabricated sections—complete with installed equipment, piping, and cabling—are assembled like a massive puzzle. This approach enhances efficiency but requires meticulous planning and precision to ensure seamless integration. The complexity is further compounded by the need to balance structural integrity, stability, and passenger comfort, all while incorporating advanced technologies and adhering to stringent safety and environmental regulations. Overall, the production of a cruise ship is a monumental engineering endeavor, demanding coordinated efforts across multiple disciplines and a high degree of expertise.

ChatGPT 4o Research

Response:

Manufacturing Complexity of the Cruise Ship Hull

Hull Geometry & Plate Forming: The hull has complex curvature (a flared bulbous bow and contoured stern). These undevelopable double-curved surfaces require heavy plate bending or skilled line-heating to form. Flatter, singly-curved midship plates are simpler, but bow and stern sections need intensive forming and careful fairing for a smooth finish.

Structural Reinforcements & Features: Internal framing (girders, frames, stringers) supports the hull and superstructure. Multiple hull openings (e.g. bow thruster tunnels, shell doors) require reinforcement and precise fabrication to maintain strength. The flared bow likely has extra structure to withstand wave impacts. These features increase fabrication complexity during assembly.

Assembly & Construction: The hull is built in modules via block construction—prefabricated sections hoisted and welded together. Aligning large curved blocks is challenging but standard in cruise shipbuilding. Long weld seams require careful sequencing to minimize distortion. The design is feasible with panel-and-block methods but demands skilled fabrication and planning.

Aidalos

Response:

The cruise ship design presents significant production challenges, primarily due to its complex hull geometry. The bulbous bow represents the most demanding manufacturing element, featuring double curvature that requires specialized forming techniques and skilled labor—potentially requiring 10-20 times more labor hours per ton than simpler sections. The stern configuration with propeller apertures similarly demands precision manufacturing and complex welding.

While the substantial parallel midship section (stations 5-15) offers production efficiency through standardization and potential automation, the transitions between this area and the complex extremities create challenging block interfaces. Manufacturing processes are complicated by limited automation potential in curved sections, difficult access spaces, increased weld volumes, and greater distortion control challenges.

A successful production strategy must balance the efficiency advantages of the midship section against the manufacturing complexities of the bow and stern, with careful planning of block divisions, construction sequencing, and specialized fabrication approaches.

B: Check if reinforcement placement complies with general requirements and evaluate the castability. Can you suggest improvements?

https://storage.googleapis.com/agent_image_store/armering.jpg

<div>ChatGPT 4o</div> <div>Response:</div> <div>Based on the reinforcement image, here is an assessment and suggestions in approximately 150 words:</div> <div>The reinforcement placement generally aligns with standard practices, including sufficient anchorage lengths and clear cover from formwork. However, several aspects could be improved to enhance both compliance and castability. Firstly, bar spacing appears inconsistent—potentially hindering concrete flow and compaction, which increases the risk of voids or honeycombing. Spacers and chairs seem insufficient or unevenly placed, which might compromise cover uniformity. Overlaps (lap joints) should be tied and staggered properly to avoid congestion and ensure force transfer.</div> <div>Castability could be improved by simplifying reinforcement arrangements in congested areas and ensuring that all bars are adequately secured against movement during casting. It's also essential to allow sufficient clearance for vibrators to ensure full consolidation. Using prefabricated cages or modular reinforcement sections could streamline installation and improve quality.</div> <div>In summary, while the basic reinforcement intent is met, refinements in bar layout, spacing, and support would significantly enhance performance and castability.</div> <div>4/29/2025</div>	<div>ChatGPT 4o Research</div> <div>Response:</div> <div>Issues Observed</div> <div>Spacing & Cover: Some bars are placed too tightly (clear gaps under ~25 mm) and too near the formwork, violating minimum spacing and concrete cover requirements. This can trap coarse aggregate (leaving mostly paste between bars) and undermines durability.</div> <div>Bar Congestion: Heavy bar congestion (many bars overlapped in one zone) hinders proper concrete placement and compaction, raising the risk of voids or honeycombing. Poor bar alignment or excessive lap splices in one spot further worsen this issue.</div> <div>Recommended Improvements</div> <div>Ensure the code-required minimum clear spacing between bars (≈ one bar diameter or 25 mm) to allow proper concrete flow.</div> <div>Use spacers/chairs for adequate cover, and keep bars aligned to preserve spacing during the pour.</div> <div>Stagger lap splices to avoid too many overlaps in one area.</div> <div>If congestion is severe, consider using fewer, larger-diameter bars or a single rebar layer to ease casting.</div> <div>Aidalos</div>	<div>Aidalos</div> <div>Response:</div> <div>Based on my analysis and the structural specialist's assessment, here's my evaluation of the reinforcement placement:</div> <div>Reinforcement Compliance and Castability Evaluation</div> <div>The reinforcement design shows several compliance concerns and significant castability issues. The most critical problem is severe congestion at the wall-footing junction, where bundled Ø16-G1208 bars (34 total) create an extremely dense reinforcement cage in the lowest 600mm of the wall. This congestion will impede concrete flow, increase the risk of voids and honeycombing, and make proper vibration nearly impossible.</div> <div>Recommended Improvements:</div> <div>Replace bundled Ø16 bars with fewer larger-diameter bars (Ø20/Ø25) to reduce congestion while maintaining structural capacity</div> <div>Stagger the 90-degree hooks at different elevations within the footing</div> <div>Consider a haunch or thickened section at the wall base to provide more space for reinforcement</div> <div>Specify smaller aggregate size for the congested junction area to improve flowability</div> <div>Add detailed construction sequence notes for concrete placement and vibration techniques</div> <div>These modifications would significantly enhance constructability while maintaining structural integrity, ensuring proper concrete consolidation and reducing the risk of future structural defects.</div> <div>6</div>
--	--	--

C: Analyse the photo and give an assessment. What should be done?

https://storage.googleapis.com/agent_image_store/A_NB_39_photo_walls_inside.png

<div>ChatGPT 4o</div> <div>Response:</div> <div><p>The photo shows an interior wall with visible signs of water damage—staining, possible mold, and deterioration of wall materials. This suggests a leak or prolonged exposure to moisture, likely from plumbing, roof infiltration, or condensation issues. The condition may pose health risks and indicates a structural concern.</p><p>Immediate action should include identifying and fixing the moisture source. Remove damaged wall sections, dry the area thoroughly, and treat for mold if present. Replace insulation and wallboard, then repaint with mold-resistant materials. A professional assessment is recommended to ensure all damage is addressed and future issues are prevented.</p></div>	<div>ChatGPT 4o Research</div> <div>Response:</div> <div><p>The interior wall appears structurally sound, with a smooth surface and consistent construction quality; edges and corners are well finished. The finish is a painted surface that looks intact without major cracks or holes, though minor hairline cracks or scuff marks are visible. There are no obvious stains or damp patches, indicating no current water damage issues; the surface seems reasonably clean aside from light dust or minor smudges. The paint appears slightly worn or uneven in places, suggesting normal wear. Overall, there are no immediate safety hazards observed; any issues are primarily cosmetic and related to normal aging.</p><div>Recommendations</div><div><p>Clean and dust: Remove any dust and surface grime by gently wiping down the wall with a soft cloth or mild cleaner to improve appearance and prepare for repairs or painting.</p><p>Repair minor damage: Fill in the minor hairline cracks or small holes with an appropriate wall filler or spackling compound, then sand the area smooth after it dries to restore a uniform surface.</p><p>Repaint the finish: Apply a fresh coat of paint (or touch up worn sections) to cover scuff marks and uneven color, which will refresh the wall's appearance and provide a protective, uniform finish.</p><p>Monitor the condition: Regularly check the wall for any new cracks, stains, or signs of moisture over time. Address emerging issues early with maintenance or repairs to preserve the wall's integrity and appearance.</p></div></div>	<div>Aidalos</div> <div>Response:</div> <div><p>Based on my analysis of the image, I can provide this assessment:</p><p>The photo shows a building interior in early renovation stages with exposed brick and timber walls. The structure has been stripped to its core elements with new timber framing installed around a window and doorway. The dark timber walls require thorough inspection for potential rot, insect damage, or fire damage.</p><p>Next steps should include:</p><div><p>Structural assessment of the aged timber</p><p>Implementation of proper insulation solutions</p><p>Addressing potential moisture issues</p><p>Installing modern electrical and plumbing systems</p><p>Applying appropriate finishes while preserving historical elements</p></div><p>This renovation requires careful balance between modernization and preserving the building's character.</p></div>
--	--	---

D: Where should we put sensors on to monitor this bridge?

https://storage.googleapis.com/agent_image_store/pdf_conversions/KolomoenBru/AsBuilt/page_0.png,
https://storage.googleapis.com/agent_image_store/pdf_conversions/KolomoenBru/AsBuilt/page_1.png

ChatGPT 4o

Response:

To effectively monitor the Kolomoen Bridge, sensors should be strategically placed on key structural elements to detect stress, movement, and degradation. Install strain gauges and displacement sensors on the main girders and deck to track load effects and deflections. Position accelerometers at mid-span and abutments to measure vibrations and dynamic responses. Place tilt sensors on the piers and abutments to monitor settlement or shifting. Temperature and humidity sensors should be added near expansion joints and bearings to assess environmental impact. Acoustic emission or crack detection sensors can be installed near joints and known stress concentration zones for early damage detection. Additionally, install corrosion sensors within concrete near reinforcement bars, especially around the base of piers. A comprehensive monitoring system across these locations ensures early detection of issues and supports maintenance planning. Sensor placement should also consider ease of access and protection from environmental exposure for reliable long-term operation.

ChatGPT 4o Research

Response:

Sensor Placement Recommendations for Kolomoen Bridge

Piers and Abutments: Sensors on pier columns and abutment walls monitor foundation behavior, detecting settlement, tilt, or abnormal stress for early warning of support issues.

Expansion Joints: Install displacement sensors at expansion joints (located at the abutments) to track thermal movement. Monitoring joint gaps ensures the bridge expands/contracts as designed and alerts if joints seize or move excessively.

Mid-Span Girders: Place strain or deflection sensors at mid-span of main girders (where bending is highest) to measure deflection and stress, revealing any excessive response that could indicate stiffness loss or damage.

Bearings: Place sensors at support bearings to monitor movement and load distribution. This confirms bearings allow expected sliding and rotation (preventing restraint) and carry loads evenly, identifying any shifts or malfunctions.

Deck Surface: Mount accelerometers or strain gauges on the deck to monitor traffic-induced vibrations and stress, capturing dynamic responses (frequency, acceleration) that reveal changes in structural behavior or early signs of damage.

Aidalos

Response:

Based on analysis of the Kolomoen cable-stayed bridge drawings, sensors should be strategically placed on these critical structural elements:

Stay Cables: Install accelerometers at mid-length of the longest cables and strain gauges at anchorages to monitor tension forces and vibration. Focus on at least 25% of cables, prioritizing those at mid-span and quarter-span positions.

Towers: Place inclinometers and GPS receivers at tower tops, with strain gauges at tower bases and cable connection points. Add accelerometers at multiple heights to capture dynamic response.

Deck: Position displacement sensors and accelerometers at mid-span and quarter-span points. Install strain gauges at maximum moment locations and temperature sensors throughout the deck depth.

Cable-Deck Anchorages: Mount strain gauges and acoustic emission sensors around multiple anchorage points, focusing on mid-span and near-tower connections.

Foundations: Implement settlement monitoring at tower foundations and tiltmeters at foundation-tower interfaces.

This comprehensive monitoring approach enables early detection of structural issues and data-driven maintenance decisions.

E: The following is a plot from an anomaly detection campaign. Do a deep analysis and summarize the situation. What could explain the observations?

https://storage.googleapis.com/agent_image_store/a_univ_image.png

ChatGPT 4o	ChatGPT 4o Research	Aidalos
Response:	Response:	Response:
The plot shows a time series with sharp, periodic spikes in anomaly scores. Most of the time, the scores are low, but at regular intervals, they spike sharply and then decay quickly. This pattern suggests a recurring anomaly, possibly triggered by scheduled activities (e.g., batch jobs, maintenance tasks, or backups) rather than random failures. The consistent periodicity points toward a systemic cause rather than external random events. Potential explanations include scheduled data processing, nightly system reboots, or repeated sensor faults under specific conditions. Further investigation should focus on correlating spike times with known operational schedules	Observations The anomaly plot shows a mostly quiet 2019 (few, sporadic daily anomalies) compared to a volatile 2020 with more frequent and higher anomaly counts. A pronounced cluster appears in mid-2020, culminating in a single-day spike around 45 anomalies, a major outlier. Smaller surges also occur later in 2020 (several days with 20–30 anomalies), indicating multiple periods of unusual behavior rather than isolated events. Possible Explanations The mid-2020 spike and clusters might reflect a significant operational event (e.g. a system fault, maintenance, or abrupt occupancy change) causing widespread sensor deviations. Alternatively, a change in anomaly detection thresholds or sensor calibration could have inflated detections. External factors like unusual weather or a power disruption may also explain the synchronized anomaly surge. Overall, the pattern suggests a systemic disturbance rather than normal day-to-day variation.	Anomaly Detection Analysis Summary The plot shows anomaly detection data from a building control system at the University of Porto throughout 2019. The visualization consists of two charts: a stacked bar chart breaking down anomalies by metric type and a total daily anomaly count chart below it. The most significant anomaly event occurred in late May/early June, with daily counts peaking near 40. Other notable clusters appear in early September and late November/early December, with counts reaching 30-35 anomalies per day. The system monitors multiple metrics including total energy consumption/generation, outdoor temperature, and zone-specific HVAC, lighting, socket usage, and temperature readings across five zones. Possible explanations include: System malfunctions or maintenance activities Seasonal patterns affecting HVAC performance during temperature transitions Equipment issues or sensor glitches External events like power fluctuations The clustering pattern suggests cascading effects where problems in one system trigger anomalies in related systems.