AI IN AVIATION MAINTENANCE · A TECHNICAL PERSPECTIVE

The Safest AI in the Hangar Knows What It's Allowed to Do

Certified technicians don't improvise. They work from approved data, cite the manual, and sign for every task. AI in the hangar should meet the same standard.

Christopher Wells, PhD
Principal AI Scientist, AIXI

INSIDE

The alignment question
Four guardrails
AutoCoder: bounded by design
Ask OTTO: judgment within guardrails
Prescriptive Maintenance: history in view
Five questions to ask any AI vendor

AIXI

01

The Alignment Question

The AI industry is chasing an answer to one question: how do you make sure a machine makes the same call a trusted person would? Aviation maintenance has been answering that question for decades. Not for machines. For people.

The safest and most trusted AI isn't the one that can do the most. It's the one that knows exactly what it's allowed to do.

The AI industry has a name for this: alignment. Put simply, alignment is when an AI system's behavior matches the intent of the people who rely on it. Alignment asks whether the system will make the same choice a trusted expert would, in the same situation, every time.

Much of the public conversation about alignment centers on distant, hypothetical risks. In the hangar, the problem is concrete and immediate. A wrong ATA code skews the reliability data an airline depends on. A confident answer with no source behind it can send a technician down the wrong troubleshooting path on an aircraft scheduled to fly in the morning.

Business leaders already feel this. Those concerns don't call for less AI. They call for AI built like everything else in aviation: bounded, traceable, and accountable to the people who are responsible for the work.

WHAT EXECUTIVES SAY

40%

worry about AI giving incorrect information

32%

want more transparency about how their data is used

Source: "AI Adoption Is Outrunning Trust. Explainability Closes The Gap," Forbes, September 29, 2026

In aviation maintenance, an AI mistake isn't an inconvenience. It impacts airworthiness and the bottom line.

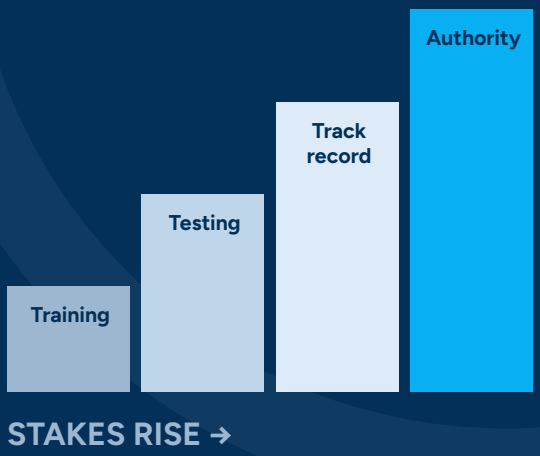
02

Earning a Place in the Hangar

So how do you build AI that meets that standard? The same way aviation builds trust in people.

Not every technician can sign off every task. Authority follows qualification. Qualification follows training, testing, and a record of getting it right. The rules tighten as the stakes rise.

AI should earn its place the same way. At AIXI, four guardrails hold across everything we build.



THE FOUR GUARDRAILS

1 **It knows its limits.** Every output is measured for confidence. When confidence is low, the system flags the result or declines to answer, rather than presenting a guess as fact.

2 **It shows its work.** Every output ties back to its source, whether that's the maintenance manual, the parts catalog, or the airline's own records and data.

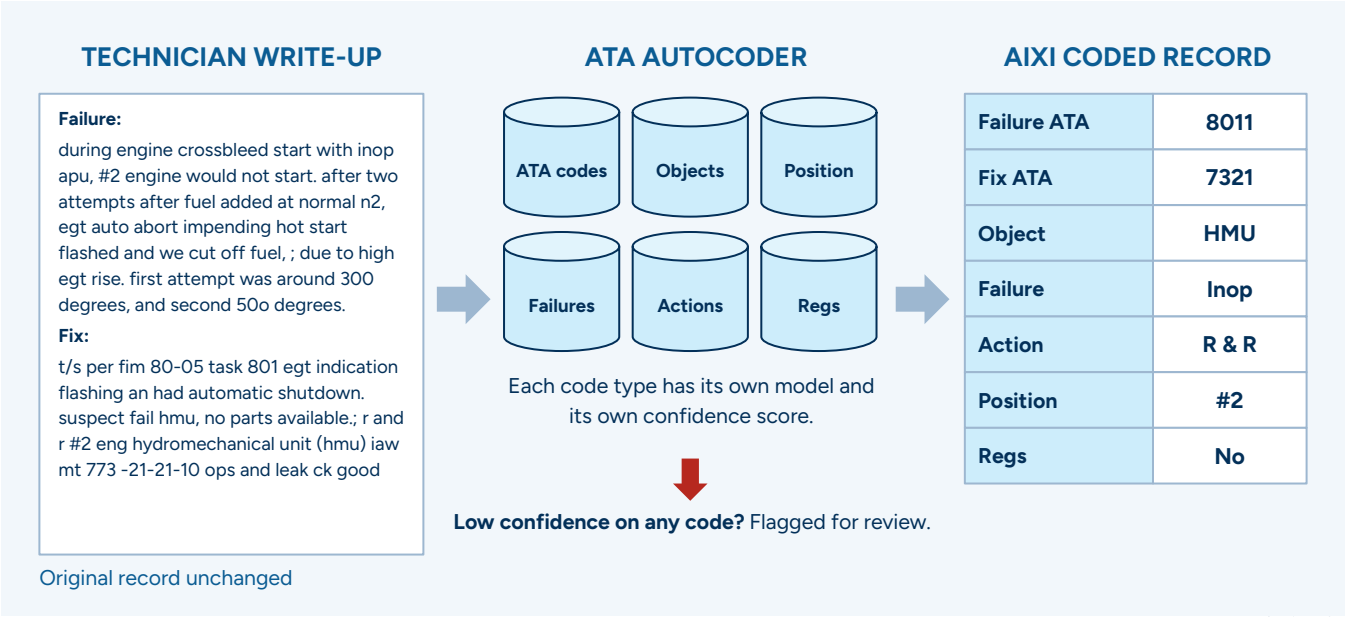
3 **It's tested, not assumed.** Model performance is measured against each airline's own thresholds, then monitored over time. When fleets or write-ups change, our model operations catch the drift and correct for it.

4 **People stay accountable.** Technicians and engineers make every airworthiness call. Reliability leaders own the business decisions. AIXI's tools surface the patterns, and people choose how to act on them.

We build two very different kinds of AI on that foundation, and how they work depends on the task.

03

The AutoCoder: Bounded by Design



Every day, technicians write thousands of maintenance entries in their own words: shorthand, abbreviations, misspellings typed at the end of a long shift. That free text holds a fleet's maintenance history, but it only becomes useful once each entry carries an appropriately assigned ATA code.

That's the job of AIXI's ATA AutoCoder. It reads each entry exactly as written and assigns the code in real time. It doesn't alter the original record. The technician's words stay the technician's words.

Here, the alignment question is narrow by design. The ATA AutoCoder can only choose from ATA codes defined in the aircraft's manual or already seen in the maintenance data. It can't invent a new one. When it isn't confident, it says so, and the entry is flagged for a person to review.

Each airline's experts review and correct the ATA AutoCoder's early results, building a truth set that fine-tunes the models to their standards. Once that foundation is in place, only about 5% of write-ups need human review, so experts spend their time where it counts.

And since our models apply the same logic to every entry, the ATA AutoCoder can be up to nine times more consistent than human coders. The same fault gets the same code, no matter who wrote it up or when.

Up to 9x

more consistent than human coders

AIXI analysis of 142,586 held-out maintenance records

The result isn't just efficiency. It's reliability data an airline can stand behind.

04

Ask OTTO: Judgment Within Guardrails

Every maintenance operation runs on hunches. We've all heard statements like, "It feels like we've had more parking brake problems lately," or "It seems like one station keeps writing up aux pitot heat failures." Feelings come through in meetings, and sometimes they drive decisions.

Ask OTTO puts assumptions to the test. A reliability engineer asks, in plain language, for every aux pitot heat failure over the past twelve months, grouped by station and fleet. Ask OTTO answers from the airline's own records: five reports at five different stations, one each, all on the same fleet type. One station didn't stand out. The pattern everyone felt wasn't there. The supporting data sits right below the answer, so the engineer can check it. Data, not gut feel.

SAMPLE QUERY FROM ASK OTTO

You
report the most common discrepancy by aircraft number by fleet on Fridays for the past 4 months.

Summary: The query returns a list of the most frequently reported issues (referred to as 'discrepancy') for each aircraft, identified by their 'acn' (aircraft number), and grouped by fleet type ('fleet'). The reported issues occurred only on Fridays during the last four months. The 'open_date' column was not present in the final output as it was not part of the groupby operation used to generate the results. The 'acn' column represents the identifier for each specific aircraft. The 'fleet' column indicates the type of fleet the aircraft belongs to, and the 'discrepancy' column provides the details of the issues reported for each corresponding aircraft number.

acn	fleet	discrepancy
950	ng	crew o2 low ;
486	ng	22 def seat legs loose at track ;
439	ng	fsd, fwd lower corner area has pressure leak. ;
8885	max	fwd coffee maker leaking ;
441	ng	opened #1 eng fan cowls to f.o.m ;

This is where alignment gets harder.

Unlike the ATA AutoCoder, Ask OTTO makes judgment calls. It interprets the question and decides which records matter. So we built firm rules around it:

1. **It only knows what you give it.** Ask OTTO works from the data the airline chooses to share. Nothing more.

2. **It shows the evidence.** Every answer comes with the records behind it.

3. **It says when it can't answer.** If the data isn't there, Ask OTTO says so rather than guessing. The airline decides whether to share more information to answer more questions.

4. **It points to the source.** When a technician asks about a fault or a part, Ask OTTO shows where the answer lives in the maintenance manual or the parts catalog. It doesn't tell them what to do.

People stay in command. Ask OTTO gets them to the facts sooner, but it never replaces their judgment.

05

Prescriptive Maintenance: History in View, Decision in Hand

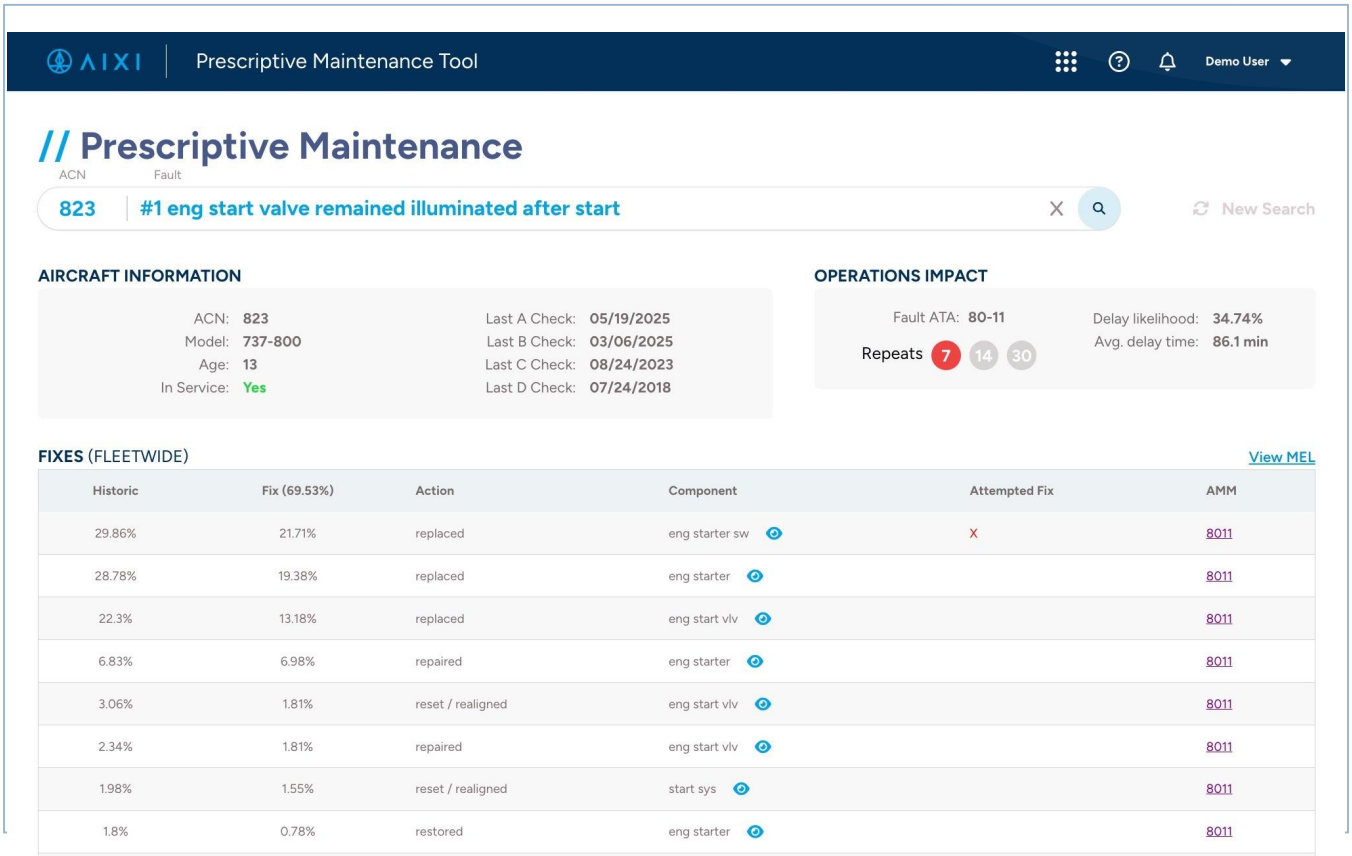
When a fault comes in, a technician's best resource is often the last person who fixed it. That knowledge sits in years of write-ups, but few people have time to dig for it.

AIXI's Prescriptive Maintenance brings that history forward. For a recurring fault, it shows what has been tried before, which fixes held, and which didn't, all drawn from the airline's own consistently coded records.

Every recommendation links back to the write-ups behind it.

The guardrail is simple: it informs the decision; it doesn't make it. The technician sees the evidence, applies the manual, and signs for the work. The system's job is to make sure no one starts from a blank page.

SAMPLE FAULT IN THE PRESCRIPTIVE MAINTENANCE TOOL



Historic	Fix (69.53%)	Action	Component	Attempted Fix	AMM
29.86%	21.71%	replaced	eng starter sw	X	8011
28.78%	19.38%	replaced	eng starter		8011
22.3%	13.18%	replaced	eng start vlv		8011
6.83%	6.98%	repaired	eng starter		8011
3.06%	1.81%	reset / realigned	eng start vlv		8011
2.34%	1.81%	repaired	eng start vlv		8011
1.98%	1.55%	reset / realigned	start sys		8011
1.8%	0.78%	restored	eng starter		8011

Fleetwide fix history for one fault, with the AMM reference on every row.



06

Built the Way Aviation Already Works

None of this is new thinking for aviation. It's how the industry has always brought in new technology, and regulators are applying the same logic to AI.

FAA[1]

The FAA's roadmap for AI safety assurance asks that AI fit within existing safety processes rather than sit outside them. It warns against treating AI as if it were a person, because accountability must stay with people. It calls for starting with lower-risk applications and names predictive maintenance as a research priority.

EASA[2]

EASA's guidance for machine learning in aviation points the same way:

- Quality data,
- Explainable outputs, and
- Meaningful human oversight.

Every guardrail that AIXI uses traces back to one of those principles. Our tools work inside the maintenance processes airlines already run. They inform people. They don't take over airworthiness decisions. And they've earned their place gradually, in production at airlines since 2022.

That isn't compliance for its own sake. The CEO of the American Institute of Aeronautics and Astronautics (AIAA) recently asked whether aerospace can move as fast as the moment requires without sacrificing safety.[3] At AIXI, guardrails are how you get both.

1 FAA, Roadmap for Artificial Intelligence Safety Assurance, Version I, 2024.
2 EASA, Artificial Intelligence Concept Paper, Issue 2, 2024.
3 C. Mowry, "The Conversations Aerospace Needs Right Now," Aerospace America (AIAA), Oct. 1, 2026.

07

Five Questions to Ask Any AI Vendor

The broader AI industry will keep debating alignment. In the hangar, the answer is already written into how the work gets done: approved data, cited sources, and a signature on every task. AI should be held to that same standard.

If you're evaluating AI for maintenance, reliability, or airworthiness, start with five questions:

1. **What data does it work from, and who controls it?**
2. **Can it show the evidence behind every answer?**
3. **What does it do when it isn't sure?**
4. **Who stays accountable for the decision?**
5. **How is it tested, and how will you know it still performs next year?**

A strong AI or technology partner will answer each one plainly.

A vague answer is an answer too.



Christopher Wells, PhD

Principal AI Scientist at AIXI, where he leads the technical architecture of an AI platform built for aviation maintenance and reliability. A physicist turned AI engineer, he holds a PhD in high-energy theoretical physics from Johns Hopkins University and has spent more than 15 years building production machine learning and generative AI systems for regulated, mission-critical environments. Chris's work spans LLM orchestration, multi-agent systems, RAG pipelines, and AI governance.

If you're weighing how AI fits into your maintenance and reliability operation, we'd welcome the conversation. Contact John Bowell to kick off a dialogue or see a demo.

JBowell@aixi.co